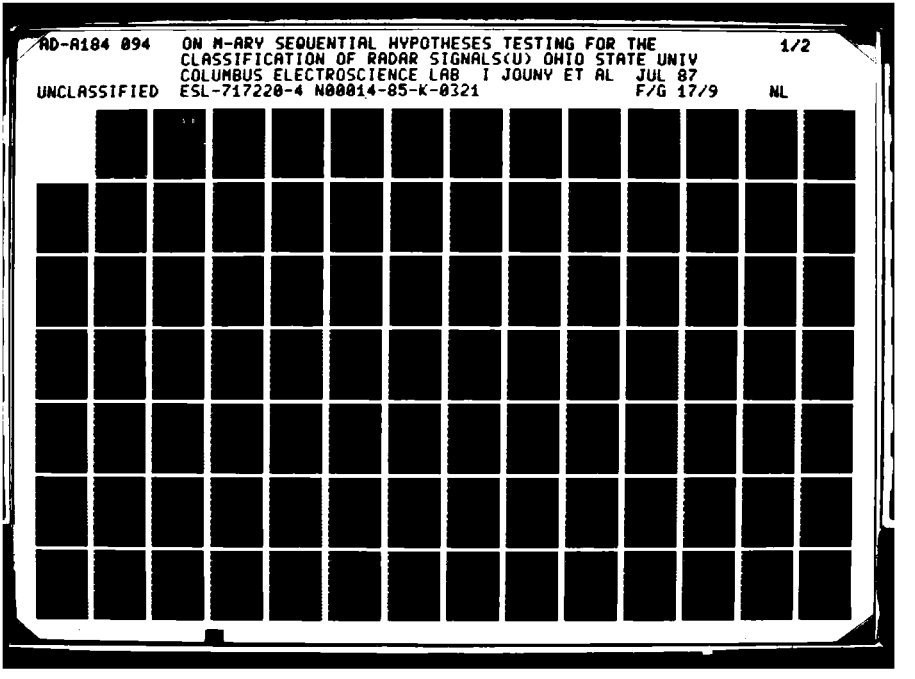
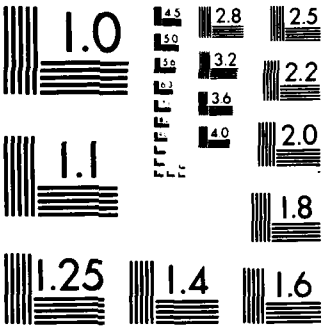

AD-A184 894
ON M-ARY SEQUENTIAL HYPOTHESES TESTING FOR THE CLASSIFICATION OF RADAR SIGNALS(U) OHIO STATE UNIV COLUMBUS ELECTROSCIENCE LAB I JOUNY ET AL JUL 87
1/2
UNCLASSIFIED
ESL-717220-4 N00014-85-K-0321
F/G 17/9
NL



THE OHIO STATE UNIVERSITY

ON M-ARY SEQUENTIAL HYPOTHESES TESTING FOR THE CLASSIFICATION OF RADAR SIGNALS

Ismail Jouny
F.D. Garber

The Ohio State University
ElectroScience Laboratory
Department of Electrical Engineering
Columbus, Ohio 43212

Technical Report No. 717220-4
Contract No. N00014-85-K-0321
July 1987

Department of the Navy
Office of Naval Research
800 N. Quincy Street
Arlington, Virginia 22217

NOTICES

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely related Government procurement operation, the United States Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

50272-101

REPORT DOCUMENTATION PAGE	1. REPORT NO. ADA184094	2.	3. Recipient's Accession No.
4. Title and Subtitle: ON M-ARY SEQUENTIAL HYPOTHESIS TESTING FOR THE CLASSIFICATION OF RADAR SIGNALS			5. Report Date: July 1987 6.
7. Author(s): Ismail Jouny and F.D. Garber			8. Performing Organization Rept. No: 717220-4
9. Performing Organization Name and Address: The Ohio State University ElectroScience Laboratory 1320 Kinnear Road Columbus, Ohio 43212			10. Project/Task/Work Unit No. 11. Contract(C) or Grant(G) No (C) N00014-85-K-0321 (G)
12. Sponsoring Organization Name and Address: Department of the Navy Office of Naval Research 800 N. Quincy Street Arlington, Virginia 22217-5000			13. Type of Report & Period Covered: Technical 14.

15. Supplementary Notes

16. Abstract (Limit: 200 words)

This report is concerned with the performance of M-ary sequential hypothesis tests applied to the classification of radar signals. In the first part of this report, several available M-ary sequential algorithms are considered. These include the algorithms due to Reed, Armitage and Palmer. Modifications of the existing algorithms are considered and their effects on the average number of required measurements and the resulting error performance are examined. Three other techniques are also proposed, a tree algorithm and a sequential maximum a posteriori test as well as a sequential version of the nearest neighbor. The second portion of the report is devoted to the performance evaluation of the various algorithms by means of Monte-Carlo simulations of the radar signal observations. The performance of these techniques is compared and the relative merits of each algorithm are discussed.

17. Document Analysis a. Descriptors

b. Identifiers/Open-Ended Terms

c. COSATI Field/Group

18. Availability Statement	19. Security Class (This Report)	21. No of Pages
Approved for public release; distribution is unlimited.	Unclassified	104
	20. Security Class (This Page): Unclassified	22. Price

(See ANSI-Z39.18) See Instructions on Reverse OPTIONAL FORM 272 (4-77) (Formerly NTIS-35) Department of Commerce

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Classification of an unknown target from radar return signals by means of sequential hypothesis testing techniques is the subject of this study. A radar target identification (RTI) system is to be designed to distinguish measurements of the radar backscatter from an unknown object as belonging to one of a set of M classes, with each class corresponding to a particular airborne radar target.

The classification of targets observed by a radar or other sensor is of prime importance in ballistic missile defense and other similar problems. Targets are (at least conceptually) examined one at a time and classified based on the time ordered returns from a set of consecutively transmitted pulses. In this report an observation is considered to be a sample waveform from a random process, and the classification algorithm or *classifier* is based on sequential statistical hypothesis testing.

The radar system to be used is a general purpose, multifrequency, multipolarization system. It operates in the range from 8 to 58 MHz in horizontal receive mode (see[1]). For the targets of interest, these frequencies represent the resonant region of a catalogue of radar targets which are used for the experimental phases of the study [2]. The resonant region corresponds to the band of frequencies with wavelengths which are approximately equal to the dimension of the target.

A reasonable design goal for a target identification system is to realize an al-

gorithm that is capable of producing a reliable decision with as few measurements as possible. It has long been recognized [3,4] that sequential hypothesis testing techniques provide a reasonable compromise in the tradeoff between the classification error rate and the average number of measurements, $E\{n\}$, required to reach a decision.

The theory of sequential hypothesis testing was developed for the binary (two hypotheses) case by Wald [3]. Since that time, this theory has had a wide variety of applications. For the application to target identification, the target returns are observed in stages. At each stage a decision is made either to classify the target (as a particular object type) or to make another observation.

For the binary case, the sequential classification procedure is optimal in the sense that a classification is made and the decision sequence ends with the minimum number of returns necessary to achieve a prescribed probability of error [5]. Sequential classifiers provide important advantages over those classifiers that employ a predetermined, fixed number of return measurements. Targets that are easy to identify are classified quickly, while targets that are more difficult to identify can be observed for a longer period of time before reaching a decision. This results in a more efficient use of the sensor and of computational resources and an overall improvement in the classification performance. Although most of the important results from the theory of sequential testing do not require statistical independence between successive observations, independence of observations greatly simplifies the design and analysis of the performance of the sequential test.

Since the original work of Wald [3], a number of techniques have been proposed that extend binary hypothesis testing methods to the case of $M \geq 3$ alternatives [6,7,4,8]. Each of these techniques realize some performance characteristic that may be desirable for certain applications. In addition to the classification error

rate and the average number of measurements, these characteristics may include the maximum number of measurements allowed, the complexity of implementation, and the sensitivity to noise power levels. The relative importance of these performance parameters depends on the application.

The present consideration of M-ary sequential techniques is primarily motivated by interest in the reliable identification of aircraft. For this particular application, each measurement X is a vector whose component x_i, $i = 1, \ldots, K$ are complex numbers representing the in-phase and quadrature parts of the backscatter signal at a particular frequency, f_i.

The target, whose identity is unknown, may be at unknown azimuth and elevation relative to the radar. In cases where the aspect angle (azimuth and elevation) of the object are known, the M hypotheses may be regarded as "simple"; each hypothesis corresponding to a target class described by a single prototype. When the aspect angle of the object is unknown, or known to be within some range of angles, the decision must be made among M composite hypotheses; each hypothesis corresponding to a target class containing N_s prototypes that represent the possible aspect angles in the specified range.

For experimentation purposes, simulated radar returns were obtained from the Ohio State University compact radar range as discussed in [9]. The compact range data has been normalized so that all system related parameters have been removed from the measurements. The compact range data is in units of dBm^2 which is the radar cross section of the target, relative to 1 square meter. This unit of measurement is also used to describe the average power of the noise used in the simulation.

Figure 1 shows the structure of a classifier designed to identify radar returns where observations are taken sequentially. The classifier includes three main steps:

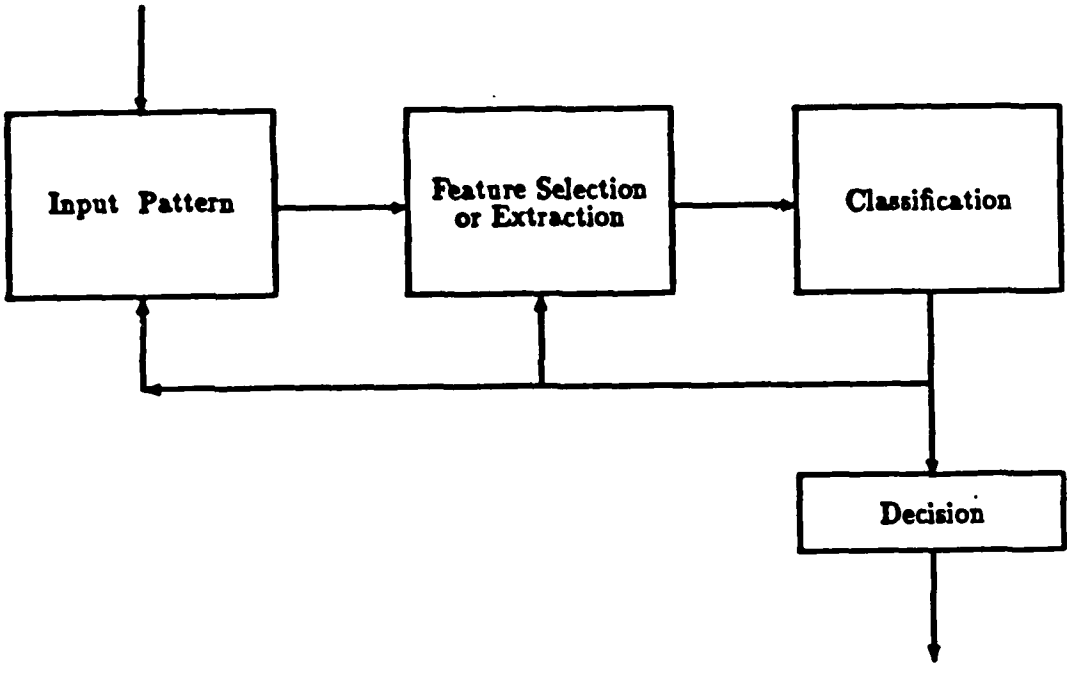


Figure 1: Sequential Pattern Recognition System

test pattern reception, feature selection and extraction, and classification. This study is mainly concerned with the classification operations to be performed to reach a decision.

Choosing an optimum set of features from the input pattern results in a reduction of both the probability of misclassification and the required number of measurements [4]. Feature ordering helps improve the performance of the classifier as a whole [4]. In this study the term "feature" refers to the frequency component

used. This means that changing features at each stage of the sequential test is a change in the frequency elements measured whenever a new observation is requested. Thus, feature ordering for radar target identification is done by selecting an optimum set of frequencies at each stage of the sequential test. Vectors of observations that correspond to different set of frequencies at each stage of the sequential test describe the target in a better way than those utilizing the same frequencies through the entire test [10]. If the sequential test involves class rejection, then the optimum set of frequencies depends on the nature and number of classes that are not rejected. The optimum set of frequencies include those that give the most "recognizable" returns when applied to different type of targets.

In this study, the set of frequencies used is an optimum set obtained by application of feature selection algorithms [10]. The sequential observations are in fact utilizing repeated measurements of the same set of optimum frequencies, but the noise content of each measurement differs from stage to stage. Repeating the same observations many times results in an effective improvement in the signal to noise ratio [11]. As more observations are taken and the classifier makes use of the previous measurements, the test eventually reaches a stage where a reliable decision can be declared.

1.1 Purpose of the Study

The optimal classifier is to be designed such that it requires the least possible number of observations given a fixed probability of misclassification. To achieve this goal, an investigation of the algorithms that have already been proposed in literature is carried out and a comparison between these algorithms is made based on their performance.

Modifications of the techniques are suggested as a measure to improve their

performance tradeoff between the expected number of observations $E\{n\}$ and the probability of misclassification. Ways of converting some techniques that require a predetermined fixed number of observations into sequential techniques are also suggested in this study.

This study considers the case where complete a *priori* knowledge of the statistics of the random observations is available and the case where such a knowledge is not complete. The former case is known as "parametric" and the latter as "nonparametric".

The concept of sequential hypotheses test for the binary case, as introduced by Wald, is discussed in Chapter II. Several generalizations of this technique for the M-ary case and some modifications of these techniques are discussed in Chapter III. An M-ary sequential classification technique based on a tree structured algorithm is proposed in Chapter III. In Chapter IV, a sequential version of the nonparametric nearest-neighbor (NN) method of pattern recognition is considered.

The performance of the various techniques is evaluated in Chapter V by means of computer simulation studies. In obtaining the results for Chapter V, the radar signals are simulated using a set of stored reference patterns of five different commercial aircraft, each corresponding to a class containing vector prototypes representing observations of a particular aircraft at up to nineteen different azimuth angles. (See [1] for a discussion of the generation and characteristics of the aircraft catalog database.)

CHAPTER II

The Concept of Sequential Hypotheses Testing

A sequential test is an adaptive procedure to decide among two or more alternate hypotheses, where observations are taken sequentially until a decision is available. If the parameter to be minimized is the average number of observations (samples), then Wald [3] showed that the sequential probability ratio test (SPRT) for the binary case (two hypotheses) requires in the mean the least number of measurements. The SPRT, being superior in this sense to the classical fixed number of observations tests, has been given much attention in the last three decades. The sequential probability ratio test, as a binary hypotheses test, has become very important in the field of radar detection. The application of sequential tests to the detection problem minimizes the average detection time.

2.1 Observations and Associated Probability Distributions

Consider the problem of classifying a set of n, K-dimensional vector observations, $X^{(n)} = \{X^1, X^2, \ldots, X^n\}$, $X^l \in C^K$ as belonging to one of M classes, where each class corresponds to an event, ω_i for $i = 1, \ldots, M$. Let $p_n(x/\omega_i)$ denote the joint conditional density function of the n random K-dimensional vector observations. Let $P(\omega_i)$ denote the a *priori* class probability for the event ω_i, $i = 1, \ldots, M$. We assume that class ω_i is composed of N_s subclasses, corresponding to the events $\omega_{i,1}, \omega_{i,2}, \ldots, \omega_{i,N_s}$ each of which represents the target of

class ω_i at a distinct azimuth angle, and let $p(x^l/\omega_{i,j})$ denote the joint conditional probability density function of the l^{th} observation $X^{(l)} = x^{(l)}$ given subclass $\omega_{i,j}$.

The *a posteriori* probability of the event (class) ω_i is given as

$$P(\omega_i/x^l) = \frac{P(\omega_i)p(x^l/\omega_i)}{p(x^l)} \tag{2.1}$$

where

$$p(x^l/\omega_i) = \sum_{j=1}^{N_s} P(\omega_{i,j})p(x^l/\omega_{i,j}) \tag{2.2}$$

Assuming the M classes are equally probable, then for Gaussian distributed random observations the joint conditional density function, which is a Gaussian mixture (for n observations), is expressed as follows:

$$p_n(x/\omega_i) = \prod_{l=1}^{n} p(x^l/\omega_i) \tag{2.3}$$

$$= \prod_{l=1}^{n} \frac{1}{N_s} \sum_{j=1}^{N_s} \frac{1}{(2\pi)^{\frac{K}{2}}\sigma^K} \exp\left[\frac{-\sum_{k=1}^{K} Re^2\{x_k{}^l - s_{i,j,k}\} + Im^2\{x_k{}^l - s_{i,j,k}\}}{2\sigma^2}\right] \tag{2.4}$$

Where x_k^l is the k^{th} frequency component of the l^{th} measurement vector, and $s_{i,j,k}$ is the k^{th} frequency component of the prototype corresponding to the j^{th} subclass (azimuth angle) of class ω_i. The joint density function, $p(x^l)$ in (2.1) is given by

$$p(x^l) = \sum_{i}^{M} P(\omega_i)p(x^l/\omega_i) \tag{2.5}$$

Let the conditional probability of deciding hypothesis ω_i when the unknown target is a member of class ω_j be $e(i,j)$. Thus, $e(i,j)$ $i \neq j$ is the probability of misclassifying a target from class ω_j, while $e(j,j)$ is the probability of correctly classifying that target.

2.2 Binary Case: The Sequential Probability Ratio Test

The sequential probability ratio test (SPRT) and other sequential hypothesis tests discussed below are based on the set of pairwise likelihood ratios, $L_{i,j}(n)$ calculated at the n^{th} stage of the test, among M possible hypotheses given as

$$L_{i,j}(n) = \frac{p_n(x/\omega_i)}{p_n(x/\omega_j)} \tag{2.6}$$

for $i, j = 1, \ldots, M$.

The sequential probability ratio test due to Wald [3] for testing a sample hypothesis or class ω_1 against a single alternative ω_2 ($M = 2$) proceeds as follows:

1. Compute the likelihood ratio $L_{1,2}(n)$ based on n observations,

2. If $L_{1,2}(n) \geq A$, decide ω_1

   If $L_{1,2}(n) \leq B$, decide ω_2

   Otherwise, increment the number of measurements, n, and repeat the test.

In this test, the parameters A and B are chosen so that:

$$A = \frac{1 - e(2,1)}{e(1,2)} \tag{2.7}$$

$$B = \frac{e(2,1)}{1 - e(1,2)}. \tag{2.8}$$

This test is optimal in the sense of minimizing the probability of error with the fewest average number of required measurements, $E\{n\}$ [3].

Equations (2.7) and (2.8) define the decision boundaries which partition the feature space into three regions: the region where ω_1 is chosen; the region where ω_2 is chosen; and the region of indifference (or null region). The null region is the region in which no terminal decision is made. This region is a major factor in determining the total number of observations and the probability of misclassification.

As the null region becomes larger, the test becomes longer and more observations must be taken, and the probability of error is reduced.

2.3 Example

Suppose that $x_1, x_2, \ldots, x_n$ are n independent measurements each with probability density function $p(x/\omega_i)$, $i = 1, 2$, a univariate Gaussian with mean s_i and variance σ^2.

The likelihood ratio $L_{1,2}(n)$ at the n^{th} stage of the test is given as:

$$\lambda_n = \log(L_{1,2}(n)) = \sum_{j=1}^{n} \log\left(\frac{p(x/\omega_1)}{p(x/\omega_2)}\right) \tag{2.9}$$

$$\lambda_n = \frac{(s_1 - s_2)}{\sigma^2} \sum_{j=1}^{n} [x_j - \frac{1}{2}(s_1 + s_2)] \tag{2.10}$$

The decision procedure becomes as follows:

$$\sum_{j=1}^{n} x_j \geq \frac{\sigma^2}{s_1 - s_2} \log A + \frac{n}{2}(s_1 + s_2) \tag{2.11}$$

$$\sum_{j=1}^{n} x_j \leq \frac{\sigma^2}{s_1 - s_2} \log B + \frac{n}{2}(s_1 + s_2) \tag{2.12}$$

Otherwise another observation is required.

The distance between thresholds that define the null region:

$$d = \frac{\sigma^2}{s_1 - s_2} \log \frac{A}{B} \tag{2.13}$$

Notice that as the variance σ^2 increases or as the class means s_1 and s_2 approach each other, the null region becomes larger. Thus, more measurements are required if the noise level increases, or if the targets to be classified are similar to each other. This is due to the fact that as the null region becomes larger, more observations

are required in order to drive the test out of this region to either of the terminal decision regions.

In the above example, notice that the decision regions of the SPRT were originally fixed by the thresholds A and B. However, these regions depend on the statistical parameters of the random observation x. At high noise levels, more observations are required before a terminal decision can be made.

The experimental phase of this study deals with the performance of a radar target identification system employing sequential techniques where the noise level affects the decision boundaries and consequently the overall error probability. If these parameters (variance or class means) are fixed then the boundaries of the decision areas are uniquely defined by the thresholds A and B and hence by the error probabilities $e(1,2)$, and $e(2,1)$.

Prediction of the number of observations required before terminating the sequential test might give an idea about the test length. In [3] an expression is given for the expected number of measurements $E\{n\}$ that a SPRT requires assuming a Gaussian distribution. Predicting the expected number of samples in the binary hypotheses tests is much simpler than that of the M-ary hypotheses tests. This is due to the fact that the probability analysis in the binary case is not as complicated as the M-ary case, especially in defining the decision regions and the corresponding error probabilities.

2.4 Modified Sequential Probability Ratio Test (MSPRT)

The SPRT due to Wald is optimal in the sense of minimizing the average number of observations $E\{n\}$ with fixed error probability. It is, however, possible for this test to require an unreasonable number of measurements before reaching a

decision. For this reason, various modifications of this test have been considered, including an abrupt truncation of the test at some value $n = N$.

As an alternative to an abrupt truncation, a modification of the thresholds, A and B is suggested in [4]. Here, the decision boundaries incorporate a dependence on n so that the new thresholds approach a common level at $n = N$ (see Figure 2). The form of the modified thresholds given in [4] is

$$A(n) = \exp\left[c_1(1 - \frac{n}{N})^r\right] \tag{2.14}$$

$$B(n) = \exp\left[c_2(1 - \frac{n}{N})^r\right] \tag{2.15}$$

for constants c_1, c_2 and r. Notice that as the maximum number of measurements, N increases, this test reduces to Wald's test.

In [4] it is shown that with a proper choice of the coefficients c_1, c_2 and r, the number of measurements can be limited while retaining the low error probability of the Wald test. The relation between the average number of measurements required for the Wald test $E_w\{n\}$ and the average number of measurements required for the modified test $E_1\{n\}$ is given by

$$E_1\{n\} = \frac{E_w\{n\}}{1+r} \leq E_w\{n\} \tag{2.16}$$

Thus, the modified test limits the maximum number of measurements and reduces the average number of measurements at the expense of a slight increase in the error probability.

Using the modification discussed above, the decision procedure in the previous example becomes as follows: if

$$\sum_{j=1}^{n} x_j \geq \frac{\sigma^2}{s_1 - s_2} c_1 \left(1 - \frac{n}{N}\right)^r + \frac{n}{2}(s_1 + s_2) \quad x \in \omega_1 \tag{2.17}$$

also, if

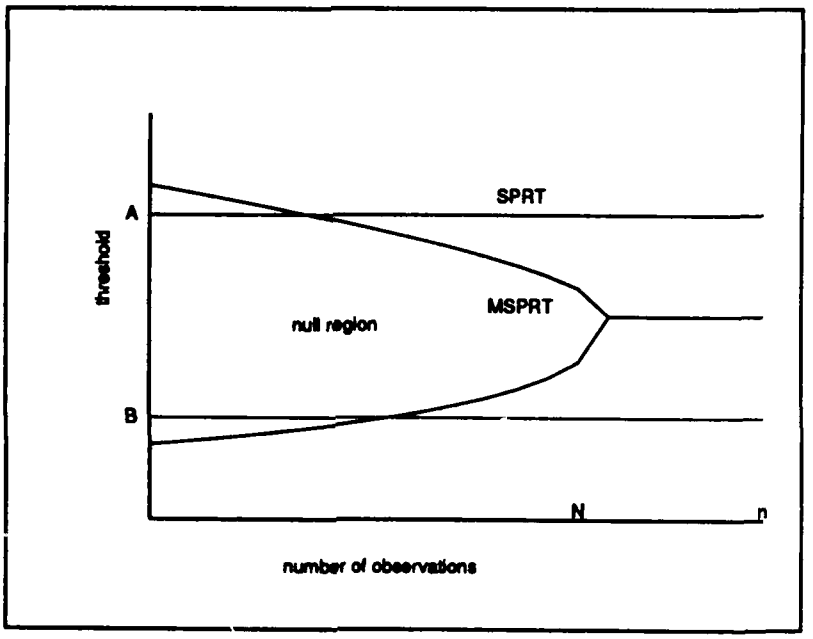


Figure 2: SPRT and MSPRT decision regions

$$\sum_{j=1}^{n} x_j \leq \frac{\sigma^2}{s_1 - s_2} c_2 \left(1 - \frac{n}{N}\right)^r + \frac{n}{2}(s_1 + s_2) \quad x \in \omega_2 \tag{2.18}$$

The distance between thresholds at the n^{th} stage of the test is

$$d_n = \frac{\sigma^2}{s_1 - s_2} \left[(c_1 - c_2)\left(1 - \frac{n}{N}\right)^r\right] \tag{2.19}$$

Notice that as $n \rightarrow N$ the separation between the two decision boundaries approaches zero. At $n = N$ the region associated with ω_1 and the region associated with ω_2 meet, eliminating the null region, and the test is terminated. In [4] it is shown that by adjusting the starting points of stopping boundaries it is possible to achieve error probabilities as nearly as low as those in Wald's SPRT.

2.5 Group Sequential Tests

In a group sequential test, observations are taken in groups rather than single observation at a time. The motivation behind this approach is the fact that the classical SPRT is a complex procedure that requires many operations at each stage of the test. A comparison must be performed after each observation, a feedback signal is required to request an additional observation, and a new hypotheses test is required at each stage of the test. The sequential observation policy suggested in [12] reduces the computation time of the SPRT and proceeds as follows:

1. At each stage of the sequential test, a test statistic is calculated based on N_0 observations.

2. A two-threshold test similar to the Wald test is performed. No more observations are requested when any of the Wald thresholds is crossed. Otherwise, this group of observations is discarded, a new group is observed and the test is repeated.

This approach is called a "Memoryless grouped-data sequential procedure". This technique is easier to implement than the Wald test, but is not optimal in any sense. However, this approach has better performance than tests with fixed number of observations [12]. The reduction in the average number of measurements using this approach can reach 60 percent compared to tests which use fixed number of observations. The SPRT, on the other hand, reduces the number of measurements up to 72 percent (see [12]). Truncation of the above test is simpler than the truncation of the SPRT because the number of observations in this case is geometrically distributed [12].

The above technique is useful whenever the number of observations required to terminate the sequential test is large. This might not be the case in radar target identification where minimizing the number of observations (test length) is the main concern.

2.6 Relative Efficiency of the Sequential Probability Ratio Test

The efficiency of the sequential probability ratio test is defined as the ratio of the expected number of measurements in the sequential test to the number of measurements required by a fixed number of observations test to achieve the same error probability. This ratio represents the performance of the sequential test in classifying targets with a certain error probability and minimum number of observations.

The relative efficiency is derived for the binary hypotheses case in [13,4]. In [13], the efficiency of the Wald sequential test to discriminate between the two hypotheses, ω_1, and, ω_2 is given as:

$$\eta = -2\frac{e(2,1)\log\left(\frac{1-e(1,2)}{e(2,1)}\right) + (1-e(2,1))\log\left(\frac{e(1,2)}{(1-e(2,1)}\right)}{(\phi^{-1}(e(2,1)) + \phi^{-1}((e(1,2))^2)} \tag{2.20}$$

when ω_2 is true, and

$$\eta = 2\frac{e(1,2)\log\left(\frac{1-e(1,2)}{e(2,1)}\right) + (1-e(1,2))\log\left(\frac{e(1,2)}{(1-e(2,1)}\right)}{(\phi^{-1}(e(2,1)) + \phi^{-1}((e(1,2))^2)} \tag{2.21}$$

when ω_1 is true, where $\phi^{-1}(\cdot)$ is the inverse of the standard normal distribution $\phi(\cdot)$. For derivation of the above equations see [13].

Notice that the efficiency of the SPRT depends on the error probabilities chosen to define the decision boundaries of the sequential test. Since no estimates for the average number of observations required by M-ary sequential tests exist, the

relative efficiency of any multiple hypotheses test can be computed experimentally only.

CHAPTER III

M-ary Hypothesis Tests: Parametric Techniques

This chapter is concerned with parametric M-ary hypotheses testing techniques where M is the number of possible classes, $M \geq 3$. The joint conditional density function $p_n(x/\omega_i)$ of the n random K-dimensional vector observations is assumed to be known. The *a priori* class probabilities $P(\omega_i)$ for the event $\omega_i \quad i = 1, \ldots, M$, are also known. Thus, the case of a random observation with known statistical parameters is considered in this chapter.

3.1 Bayes Sequential Test

From a decision-theoretic standpoint, the most reliable test for deciding among M hypotheses is the Bayes sequential procedure discussed in [4]. This test is optimal in that it minimizes the Bayes risk for a given set of cost functions and prior probabilities. Unfortunately, from the standpoint of implementation, the complexity of this test is a major concern. At each stage of the sequential process it is necessary to find the expected risk of making a decision, as well as the risk of continuing the test. Dynamic programming is used to implement this technique [4].

The intuitive argument of using dynamic programming for a finite sequential classification problem can be stated as follows: With observations taken one at a time, each stage of the test is a decision problem including both the choice of taking

additional observation or terminating the sequential test. It is easy to determine the expected risk involved in the decision when the test is terminated. However it is difficult to compute the expected risk employed in taking additional observation [14].

A new measurement is requested if its cost is less than the cost of terminating the sequential test. That is, observations are repeated if the following inequality is satisfied.

$$\begin{aligned} C(x_1,\ldots,x_n) \;+\; & \int \rho_n(x_1,\ldots,x_{n+1}) dP(x_{n+1}/x_1,\ldots,x_n) \\ & < \min_i R(x_1,\ldots,X_n;di) \end{aligned} \tag{3.1}$$

where $R(x_1,x_2,\ldots,x_n;d_i)$ is the average risk of choosing the i^{th} class after taking n measurements, $C(x_1,x_2,\ldots,x_n)$ is the cost of these n observations and $\rho_n(x_1,\ldots,x_{n+1})$ is the average risk of the $(n+1)^{th}$ observation. While it is possible to implement the Bayes procedure using dynamic programming techniques, this test has found limited applications because of the required complexity, especially in situations where rapid decisions are desired.

In the remainder of this chapter, we consider various aspects of sequential M-ary tests that are substantially less complex than the Bayes procedure. Because of the relative ease of implementation, the tests discussed below are candidate procedures for the classification of radar signals. Unfortunately, these tests produce higher error probabilities, or require more observations, on the average, than the Bayes sequential procedure. In order to compare the performance of these tests, it is necessary to evaluate the error probability and average number of required measurements for a variety of cases. The performance evaluation of these tests is the subject of Chapter V.

3.2 Pairwise Likelihood Ratios: The Armitage Test

The first M-ary technique we consider is due to Armitage [6]. At each stage of the test, this approach involves the comparison of all $M(M-1)/2$ pairwise likelihood ratios, $L_{i,j}(n)$, with a set of properly chosen thresholds, $A_{i,j}$. The Armitage algorithm is restrictive in the sense that all $M-1$ likelihood ratios for hypothesis ω_i must simultaneously exceed thier respective thresholds in order for ω_i to be selected.

This algorithm is summarized as follows:

1. Compute $L_{i,j}(n)$, $i,j = 1,2,\ldots,M$, $i \neq j$

2. If $L_{i,j}(n) \geq A_{i,j} \quad \forall j = 1,2,\ldots,M$, $j \neq i$, decide class ω_i.

   Otherwise, increment the number of measurements, n, and repeat the test.

The constants $A_{i,j}$ could conveniently be made equal to a fixed threshold A. In this case, the inequalities specify that pattern observation continues until the likelihood function of one of the hypotheses is A times each of those of the other hypotheses. In [6] it is shown that the probability of reaching a decision approaches one as the number of measurements increases. The decision probabilities, $e(i,i)$, and the thresholds, $A_{i,j}$ for this test are related by:

$$e(i,i) > 1 - \sum_{j \neq i} \frac{1}{(A_{j,i})} \tag{3.2}$$

$$A_{i,j} = \frac{\left[1 - \sum_{k \neq i} e(k,i)\right]}{e(i,j)} \tag{3.3}$$

These inequalities show that the probability of correct decision may be made large if the thresholds are chosen sufficiently large. Notice that for $M = 2$, the threshold is identical to the Wald test threshold. However, this test might be considered as

$M(M-1)/2$ binary tests where each of these binary tests has one decision area defined by the above threshold. Notice that for $M \geq 3$, the above threshold is lower than that specified by Wald for the binary case. Unfortunately, there is also a direct relationship between the size of the thresholds and the average number of required measurements. In addition, this test, like the Wald test, is not limited to a maximum number of measurements.

3.3 Modifications of the Armitage Thresholds

In order to develop a sequential test that reduces the average number of required measurements while retaining the simplicity of an approach based on a comparison of pairwise likelihood ratios, we consider a modification of the Armitage technique that is analogous to the modification of the Wald test discussed in section (2.4). In particular, we form a set of thresholds that depend on the number of measurements, n as:

$$A'_{i,j}(n) = \frac{A_{i,j}}{n^r}, \quad i = 1, 2, \ldots, M \quad i \neq j \tag{3.4}$$

where $A_{i,j}$ are the original thresholds defined by Armitage and r is a constant. Notice that if $r = 0$, the thresholds correspond to the Armitage thresholds. Also, notice that the above modification does not place a limit on the maximum number of measurements required. However, as shown in chapter V, the thresholds for the case where $r = 1$ significantly reduces the average number of required measurements, $E\{n\}$ while having little effect on the error probability of the classifier.

The motivation behind such a modification is that this technique gives very good results at low noise levels. Thus, it is sometimes possible to reduce the number of observations, provided that any slight increase in the probability of error does not change the performance of this technique as a whole. However if $r \geq 3$ then

the change in the error probability becomes significant. This is an expected result because, as r becomes significantly large, the null region is reduced at the expense of larger decision areas. Thus, terminating the sequential test becomes more likely.

The disadvantage of the above mentioned modification is that it does not guarantee a termination of the sequential test within a reasonable number of measurements.

3.4 Geometric Mean Comparison: The Reed Test

The second type of algorithm we consider is based on a comparison of the individual likelihood functions, $p_n(x/\omega_i)$ for each class, to the geometric mean of the M likelihood functions. This test, due to Reed [7], is characterized by a decision occuring on the basis of class rejection rather than class acceptence. That is, while the tests discussed above formulate a decision on the basis of the "most likely" class hypothesis, the approach used in the Reed test is a sequential elimination of the "least likely" class hypotheses until a single remaining hypothesis is chosen.

Since this method compares each of the likelihood functions to a common geometric mean, only M ratios must be computed; a significant reduction from the $M(M-1)/2$ pairwise likelihood ratios required for the Armitage test. In addition, as the least likely alternative classes are eliminated, even fewer computations are required as the test progresses until a single likelihood ratio is used to terminate the test. The implementation of the Reed approach may be summarized as in [4]:

1. For each of the M_R remaining candidate classes (initially, $M_R = M$) $\theta_i \in \{\omega_1, \ldots, \omega_M\}, i = 1, \ldots, M_R$ compute

$$U_n(x/\theta_i) = \frac{p_n(x/\theta_i)}{\left[\prod_{j=1}^{M_R} p_n(x/\theta_j)\right]^{1/M_R}} \tag{3.5}$$

2. For $i = 1, 2, \ldots, M_R$, reject class θ_i if $U_n(x/\theta_i) < A_i$.

3. If just one class θ_i remains decide hypothesis i,

   Otherwise, increment the number of measurements, n and repeat the test based on the remaining classes.

In the Reed algorithm, the thresholds, A_i are independent of the number of measurements, n and are related to the decision probabilities as:

$$A_i = \frac{1 - \epsilon(i,i)}{\left[\prod_{j=1}^{M_R}(1 - \epsilon(i,j))\right]^{1/M_R}} \tag{3.6}$$

which suggests that this test is characterized by the some of the same disadvantages as the original forms of the Wald and Armitage tests. In [4], Fu suggested the application of n-dependent thresholds, $A_i(n)$ given by

$$A_i(n) = A_i \cdot \left(1 - \frac{n}{N}\right)^r \tag{3.7}$$

for constant r, where N is the prespecified maximum number of allowed measurements.

Notice that the ratio $U_n(x/\theta_i)$ can be considered as the M^{th} root of the product of M likelihood ratios $L_{i,j}$ $j = 1, \ldots, M$ where $L_{i,j}$ is the likelihood ratio for the binary case.

The Reed algorithm requires the computation of the geometric mean of M_R hypotheses. Computational problems may result since this product is near zero whenever one of the hypotheses is unlikely. In the Armitage algorithm, such unlikely hypothesis would affect only one likelihood ratio provided that $M - 1$ likelihood functions are considered before making a decision. However in the Reed test, the unlikely hypothesis affects the entire test. Thus, whenever unlikely hypotheses exist, the decision criteria faces a numeric problem.

3.5 Modifications of the Reed Algorithm

A suggested modification of the Reed test involves changing the likelihood function $U_n(x/\theta_i)$ given above into the following form:

$$U'_n(x/\theta_i) = \frac{p_n(x/\theta_i)}{\frac{1}{M_R}\sum_{j=1}^{M_R} p_n(x/\theta_j)} \tag{3.8}$$

where $\theta_i \in \{\omega_1, \ldots, \omega_M\}$. Thus, instead of computing the geometric mean at each stage of the test, we compute the arithmetic mean. By using such a likelihood function, we eliminate the computational problems discussed above. A modification in the threshold directly follows the modification in the likelihood function. The new thresholds are:

$$A'_i = \frac{1 - e(i,i)}{\frac{1}{M_R}\sum_{j=1}^{M_R}(1 - e(i,j))} \tag{3.9}$$

Notice that if the error probabilities $e(i,j)$ are equal, then the thresholds A_i and A'_i are equal.

Using this modification a class θ_i is rejected if:

$$p_n(x/\theta_i) < A'_i \left[\frac{1}{M_R}\sum_{j=1}^{M_R} p_n(x/\theta_j)\right] \tag{3.10}$$

while, in the Reed algorithm, a class θ_i is rejected if:

$$p_n(x/\theta_i) < A_i \left[\prod_{j=1}^{M_R} p_n(x/\theta_j)\right]^{\frac{1}{M_R}} \tag{3.11}$$

The geometric mean being less than the arithmetic mean implies that the decision region for the modified test is larger than that of the Reed test. Thus, less observations are required to terminate this test. At low noise power levels the joint probability density functions $p_n(x/\theta_i)$ $i = 1, \ldots, M$ are cosiderably different from each other, thus, their geometric mean is much less than their arithmetic mean,

which means that the rejection region for the modified threshold is much larger than that defined by the Reed test. Thus, at low noise power levels, the average number of observations is reduced significantly by using the above modification.

The application of n dependent thresholds suggested by Fu [4] is still valid in this case.

3.5.1 Double Thresholds: The Reed Test

The standard Reed test employs only one decision region. However, at high noise power levels, rejecting a hypothesis is not usually done before taking many observations. Thus, a suggested modification to the Reed algorithm is to add another threshold B_i to define an acceptance region, that is: a hypothesis θ_i is accepted if $U_n(x/\theta_i) > B_i$, $\theta_i \in \{\omega_1, \ldots, \omega_M\}$. This modification improves the performance of the Reed test at high noise power levels. The threshold B_i is given as:

$$B_i = \frac{1 - \sum_{k=1, k\neq i}^{M_R} e(k,i)}{\left[\prod_{j=1}^{M_R} e(i,j)\right]^{\frac{1}{M_R}}} \tag{3.12}$$

also, for arithmetic mean comparison, the threshold B_i' is given as:

$$B_i' = \frac{1 - \sum_{k=1, k\neq i}^{M_R} e(k,i)}{\frac{1}{M_R}\sum_{j=1}^{M_R} e(i,j)} \tag{3.13}$$

Using this modification a decision can be reached even before rejecting any class. Thus, this modification reduces the hypotheses testing time in addition to reducing the average number of measurements at high noise power levels.

3.6 Single Likelihood Ratio: The Palmer Test.

Classes of sequential tests that are not direct extensions of the Wald test to M hypotheses are sequential rank tests and tests where each test statistic is compared

to multiple thresholds. Examples of tests with both of these characteristics are discussed in [15] and [16].

In [16], Palmer proposes a method based on the computation of the M likelihood functions, $p_n(x/\omega_i)$ at each stage of the test. The decisions for this test are made on the basis of the value of the ratio of the two largest likelihood functions. This single likelihood ratio is compared to a threshold, A. The implementation of this test is summarized as:

1. Compute $p_n(x/\omega_i)\ \forall\ i = 1, \ldots, M$.

2. Compute the likelihood ratio $L_{i,j}(n)$ of the largest and second largest likelihood functions.

3. If $L_{i,j}(n) \geq A$, decide ω_i,

   Otherwise, increment the number of measurements, n, and repeat the test.

In the Palmer test, the threshold A_i is given as:

$$A_i = \frac{M}{4(1 - e(i,i))^2} \tag{3.14}$$

This test gives small error probability at low noise power levels with small number of observations. However, at high noise power levels, the error probability becomes significantly large because the difference in the likelihood functions becomes very small, and a decision based on the two largest likelihood functions is not reliable. Finally, we point out that this test may also be modified to include an n-dependent threshold to allow a non-abrupt truncation while limiting the maximum number of measurements.

3.7 Sequential Maximum A Posteriori Test

The maximum *a posteriori* probability (MAP) technique chooses the hypothesis whose *a posteriori* probability is maximum. Thus, the decision is direct (nonsequential) and the target membership is assigned to the most likely hypothesis. Assuming equiprobable classes and equal cost functions, the joint conditional density functions of the n random K-dimensional vector observations are enough to decide one of the M possible hypotheses. That is, the hypothesis with largest likelihood function $p_n(x/\omega_i) \quad i = 1, \ldots, M$ is determined to be the class of the unclassified target. In this method, no thresholds are required and the decision is a one-shot procedure.

Assume n observations are taken and the sample average $\hat{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, then the expected value of $\hat{X}$, $E\{\hat{X}\} = E\{X\}$; however the variance of the sample average, $var\{\hat{X}\} = \frac{var\{X\}}{n}$. Thus, as the number of samples, n, increases, the variance of the random observation X decreases by a factor of n. This feature motivates the following algorithm:

1. Compute $p_n(x/\omega_j) \quad j = 1, 2, \ldots, M$
2. Compute $p_n(x/\omega_i) = max_{1 \leq j \leq M} \left[p_n(x/\omega_j)\right]$
3. If $p_n(x/\omega_i) \geq A_i$ Decide class ω_i,

   Otherwise increment the number of measurements, n, and repeat the test.

The test is truncated after observing a maximum number of samples N by choosing the class with maximum *a posteriori* probability. This technique gives reasonable results as shown in Chapter V.

In [17], stopping rules for this test are derived from those of the Armitage test. These stopping rules are applied to the likelihood functions rather than the likelihood ratios. Thus, equation (3.3) can be modified as:

$$\frac{p_n(X/\omega_i)}{\sum_{j=1}^{M} P(\omega_j)p_n(x/\omega_j)} \geq \frac{1}{P(\omega_i) + \sum_{j=1,j\neq i}^{M} \left[\frac{P(\omega_j)}{A_{i,j}}\right]} \tag{3.15}$$

or

$$P(\omega_i/x) \geq \frac{P(\omega_i)}{P(\omega_i) + \sum_{j=1,j\neq i}^{M} \left[\frac{P(\omega_j)}{A_{i,j}}\right]} \tag{3.16}$$

This test can be considered as a comparison of the likelihood function of a class ω_i to the arithmatic mean of all likelihood functions. Thus, it requires the computation of M likelihood ratios, while $M(M-1)/2$ likelihood ratios are required by the Armitage test. However, this test is less restrictive than the Armitage test since it does not employ pairwise comparison. Thus, the performance of this algorithm is, at the most, as good as that of the Armitage test.

This test can be modified to include hypotheses rejection (using (3.10)), where a class is rejected if:

$$\frac{p_n(x/\omega_i)}{\sum_{i=1}^{M_R} p_n(x/\omega_j)} \leq \frac{1-e(i,i)}{\sum_{i=1}^{M_R}(1-e(i,j))} \tag{3.17}$$

or

$$P(\omega_i/x) \leq \frac{1-e(i,i)}{\frac{1}{M_R}\sum_{i=1}^{M_R}(1-e(i,j))} \tag{3.18}$$

Moreover, this test can be considered as Bayes sequential test with suboptimal decision boundaries.

3.8 SPRT Applied to Tree Structured M-ary Hypothesis Tests

Algorithms for computing the optimal decision boundaries and predicting the average number of observations, $E\{n\}$, exist for various versions of sequential binary hypotheses tests. In this section the M-ary problem is treated as a sequence of binary hypotheses comparisons. The test is composed of many levels where a binary classification test is applied at each level. Thus, the M-ary test is reduced to a set of binary tests.

The radar target identification problem that requires choosing one hypothesis among M possible hypotheses can be solved by dividing the M classes into two seperate groups of classes, each including a certain set of classes that have similar features. Each of these groups is then divided into two subgroups and so on. Dividing groups into subgroups continues until two classes only are contained in each group. The Wald SPRT is then applied to various levels of the test. Thus, choosing among M hypotheses is reduced to a multilevel binary test. At each level of the test, a decision to reject a certain group or to repeat observations is made.

In order to make sure that the test terminates with a finite number of observations, truncation must be employed in the sequential decision procedure. Truncation can take place in two forms. The first form involves truncation once in the entire test (see Figure 3). If all the allowed measurements are requested before reaching the lowest level of the test i.e., before deciding between two classes, then the test continues by eliminating the null region for the rest of the binary tests. Thus, after truncation, the rest of the binary tests are considered as likelihood ratio tests.

The second form of termination employs truncation at various levels of the M-ary test. Thus, new decision regions are formed whenever a new level is reached

(see Figure 4). Truncation at each stage should be done in such a way that the total number of observations requested before the last truncation is equal to the maximum allowed number of observations. Thus, if the M-ary test consists of L levels and the maximum number of observations at each level $i = 1, 2, \ldots, L$, is N_i then $N = \sum_{i=1}^{L} N_i$ is the maximum number of allowed measurements for the entire test.

The number of levels depends on the number of classes M, as well as on the way groups are selected and the number of classes $m \leq M$, in each group. Groups may contain one or more classes depending on the way these classes are separated. A group that contains one class needs no further testing once the test reaches its level.

In this study, grouping is performed according to the physical similarities between aircraft. That is, grouping is based on physical aspects such as location of the engine, shape of the tail, size of the aircraft etc... Let $\theta_1, \ldots, \theta_m$ be the classes contained in a group Θ_i, where $\theta_j \in \{\omega_1, \ldots, \omega_M\}$ $j = 1, \ldots, m$. Also, let $p_n(x^l/\Theta_i)$ be the joint conditional density function of the l^{th} observation. Then

$$p(X/\Theta_i) = \sum_{j=1}^{m} P(\theta_j) p(X/\theta_j) \tag{3.19}$$

If n samples are observed then:

$$p(X^1, X^2, \ldots, X^n/\Theta_i) = \prod_{l=1}^{n} \sum_{j=1}^{m} p(X^l/\theta_j) P(\theta_j) \tag{3.20}$$

where $p(X^l/\theta_i)$ is given by:

$$p(X^l/\theta_i) = \frac{1}{N_s} \sum_{j=1}^{N_s} \frac{1}{(2\pi\sigma^2)^{\frac{K}{2}}} \left[\frac{\exp - \sum_{k=1}^{K} \Re^2(x_k^l - s_{i,k,j}) + \Im^2(x_k^l - s_{i,k,j})}{2\sigma^2} \right] \tag{3.21}$$

This test is not optimal because applying an optimal binary test (SPRT) at each level does not mean that the entire decision procedure is optimal. However, the

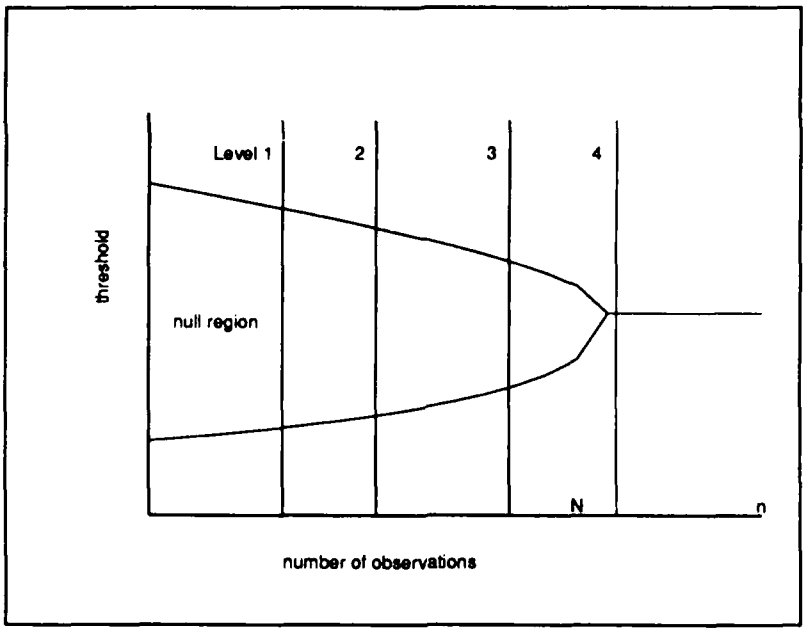


Figure 3: Decision regions dependent on the number of observations only.

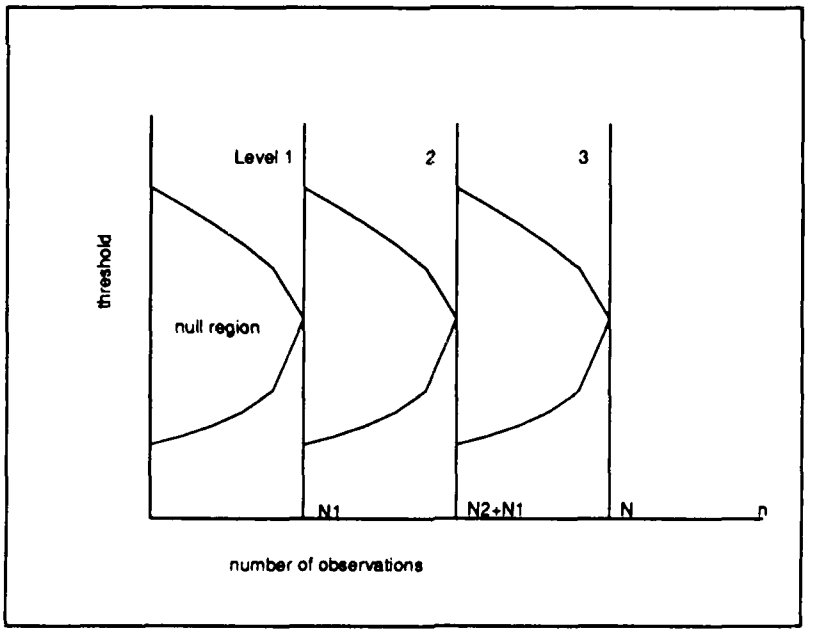


Figure 4: Decision regions dependent on the number of observations and testing stage.

average number of observations of the M-ary test which is equal to the summation of the average number of observations requested by the Wald test at each level, is predictable.

An optimal choice of groups would definitely improve the performance of the above algorithm. Moreover, the tree algorithm can be considered as a combination of both Armitage and Reed tests since it includes a pairwise comparison among groups and employs group rejection at each level of the test.

CHAPTER IV

M-ary Hypothesis Tests: Nonparametric Methods

4.1 Nonparametric Techniques

A common aspect of the algorithms discussed in the previous section is the dependence of the decision statistics on the likelihood functions, $p_n(x/\omega_i)$. Unfortunately, a useful statistical characterization of a radar system is not always available to the system designer. While statistics for the observed patterns can be estimated through learning processes, it may be undesirable to employ a parametric technique designed for a particular system in hopes that the algorithm is robust with respect to mismatched system characteristics. In this case, it is often necessary to employ some form of nonparametric classification technique. Thus, nonparametric classifiers are used for situations where the probability distribution functions of the hypotheses cannot be parametrized by a set of finite parameters. Nonparametric classifiers usually achieve good performance over a large class of distribution functions.

Fixed sample size nonparametric recognition systems have been treated extensively in the literature [11]. However, few results exist for the sequential form of nonparametric techniques. Most of the nonparametric classification techniques discussed in the literature are based on some form of rank tests [4], or on a calculation of the "distance" or least-mean-square difference between the observed signal vector, X and the set of catalog prototypes, $S_{i,j}$, $i = 1, \ldots, M$, $j = 1, \ldots, N_s$ for

the N_s prototypes of each of the M classes [8]. An example of a nonparametric technique based on a comparison of distances is the nearest neighbor (NN) algorithm which is employed in various pattern recognition applications [18,19] and is implemented in a sequential scheme later in this chapter.

4.2 Linear Sequential Pattern Classification

The linear sequential pattern classifier approach combines the sequential nature of classifiers based on sequential decision theory with the linear structure of a linear classifier [8]. In this algorithm, decisions to repeat an observation or to classify the pattern are made using linear functions derived from a set of sample patterns by the least mean-square error criterion. The decision procedure of this classifier, for a pattern X, is to measure the components $x_1, x_2, \ldots, x_K$, and classify X to class i if its image $(X \times W_n)$ lies closest to a reference point b_i; that is, if

$$||X \times W_n - b_i|| = \min_{1 \leq j \leq M} ||X \times W_n - b_j|| \tag{4.1}$$

Where the transformation matrix W_n can be expressed in terms of the sample pattern matrix S and the reference point matrix R. At the n^{th} stage of the test $W_n = S^+R$. Where S^+ is the generalized inverse of the matrix S. The decision boundaries employed are exactly the same like those defined by Armitage [6] but mapped into a decision space of the linear least mean square classifier.

The concept underlying this approach is very simple, however the implementation of this technique requires much computation at each stage of the sequential test. The implementation of this algorithm is complex because it requires the computation of the generalized inverse of the pattern matrix S whose dimensions increase whenever a new observation is requested.

In [14] a sequential classification algorithm using nonparametric ranking is proposed. This technique requires ranking of vector components that represent both the catalog data and the random observation. If single frequency radar is used and the target is of known azimuth, then this algorithm can be applied directly. However, if the observation vector utilizes more than one frequency component, then ranking is according to the norm of the vector of observations, thus resulting in a loss of information. If the azimuth position of the target is ambiguous or known only to be within a certain range then this algorithm is not useful.

4.3 Sequential Nearest Neighbor (SNN) Techniques

The nearest neighbor technique of pattern recognition is based on the computation of the vector distance between the observed signal, X and each of the class prototype vectors $S_{i,j}$ from class i, and subclass j. Specifically, the nearest neighbor algorithm decides class ω_i if $i = \arg\min_l\{\|X - S_{l,j}\|; j = 1, \ldots, N_s\}$, where $\|\cdot\|$ denotes the Euclidean distance for the K-dimensional complex vectors.

This technique may be implemented as a part of a sequential classification procedure as:

1. Compute the average, $\hat{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ of the n available observations,

2. Compute $d_{\min} = \min_l\{\|\hat{X} - S_{l,j}\|; j = 1, \ldots, N_s\}$,

$$d_{\min} = \min_l \sum_{k=1}^{K} \Re^2(x_k^l - s_{l,j,k}) + \Im^2(x_k^l - s_{l,j,k}) \tag{4.2}$$

3. Decide class ω_i if $i = \arg\min_l\{\|X - S_{l,j}\|; j = 1, \ldots, N_s\}$ and if $d_{\min} \leq A_i$. Otherwise, increment the number of observations, n, and repeat the test.

We point out that the process of averaging observations in the first step of the test may be viewed as a means to enhance the signal-to-noise ratio (SNR) of the

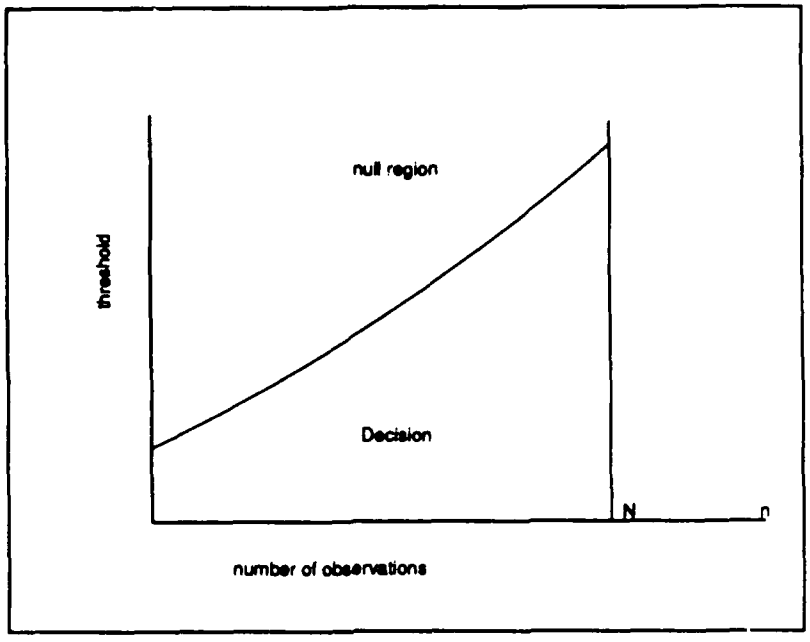


Figure 5: Decision regions for the SNN technique

decision statistic. Indeed, for coherent averaging and independent noise, the SNR at the n^{th} stage of averaging is increased by a factor of $10\log_{10}(n)$ (dB) which implies a reduction by a factor of n in the expected value of the minimum distance squared, $d_{\min}$ for the test. On the other hand, the enhancement of the SNR by averaging additional observations is directly related to the enhancement of the decision statistics based on likelihood functions for parametric tests, especially in the case of additive Gaussian noise.

In a fashion similar to that employed for the parametric tests, we may modify the thresholds in the sequential nearest neighbor test to limit the maximum number of measurements and reduce the average number of measurements while maintaining low error probabilities. In particular, we choose the measurement-dependent thresholds as:

$$A_i(n) = A_i \left(1 + \frac{n}{N}\right)^r \tag{4.3}$$

for constant r. Figure 5 shows the decision boundaries for this algorithm. Notice that the null region is above the threshold in contrast to the techniques discussed in the previous chapters. The reason for this choice of thresholds is that the sequential test becomes less restrictive as the number of observations increases.

As shown in chapter V, by increasing the value of the thresholds A_i, the classification time decreases with a slight increase in the probability of error. Thus, the performance of this algorithm, like the other parametric techniques, depends on the area of the null region, as well as the way this region varies with the number of observations.

The results of simulation studies of the performance of this nonparametric technique show that under certain circumstances, the sequential nearest neighbor test performance is comparable to that of the parametric techniques discussed above. In addition, the sequential nearest neighbor test requires only the computation of $M \cdot N_s$ vector distances at each stage of the test.

The choice of the thresholds A_i depends on the data available or the catalogue of information used for comparison. As a nonparametric approach, the nearest neighbor decision rule requires decision boundaries that are not directly related to error probabilities $e(i,j)$. The threshold A_i must be chosen such that the sequential test is not terminated before enough observations are requested. That is, no decision should be declared unless the tested target is closer to any prototype in its class than any other prototypes in the other classes. Thus, information concerning the noise power level will help in choosing optimal (or suboptimal) stopping boundaries.

An important feature of the nearest neighbor test as a nonparametric classi-

fication algorithm is that it does not require noise free catalogue data, that is the class prototype vectors can be noisy.

The above mentioned algorithm can be applied to any version of the nearest neighbor decision rule. For example, similar algorithms can be applied to the nearest neighbor with reject option. In the nearest neighbor with reject option and for every pair of integers (k, l) with $k/2 < l < k$ the k nearest neighbors of an observation are examined, and if l or more of them are in the same class the observation is assigned to this class, otherwise it is rejected.

A sequential version of the "nearest class mean classifier" can also be derived in a similar manner as the sequential nearest neighbor algorithm. The nearest class mean classifier is a classification algorithm based on choosing the class whose prototypes are closer in average to the tested target than the prototypes of all other classes. In [10] it is shown that the minimum distance classifier gives better results than the nearest neighbor method at high noise power levels. Thus, an efficient sequential classification algorithm may use the sequential version of the nearest neighbor at low noise power level and a sequential minimum distance classifier at high noise power levels. To be able to use such an algorithm, *a priori* knowledge of the noise power level should be available in advance.

CHAPTER V

Simulation Results

In this chapter, we present the results of Monte-Carlo simulation studies of the performance of the various sequential tests discussed in the previous chapters. The percentage of classification error and the average number of measurements, $E\{n\}$ are given for a number of combinations of; the total number of classes M, the number of subclasses N_s, and the dimension, K of the vector of observations, X. The maximum number of measurements is set at a nominal value of $N = 10$.

In each case, the goal is to classify an observation of an unknown radar signal as being produced by one of a set of up to five different commercial aircraft, each represented by a class containing prototypes representing vector observations of the particular aircraft at up to nineteen different azimuth angles ranging from 0° to 180°.

5.1 Database

The database consists of coherent radar backscatter measurements of scale-model of five commercial aircraft, obtained from The Ohio State University ElectroScience Laboratory compact range. The compact range data have been normalized by removing all system related parameters from the measurements. Scaled data are available for each aircraft at 0° elevation angle, and azimuth positions at 0°, 10°, 20°,..., 180°. The dimension of the vector of observations K range

from $K = 1$, to, $K = 51$, covering a frequency range 8-58 MHZ using horizontally transmitted, horizontally received polarization (HHP). Each measurement component is a complex number whose phase and amplitude are known (coherent radar backscatter). (See [1] for a discussion of the generation and characteristics of the aircraft catalog database.)

5.2 Measurements and Noise Model

The l^{th} observation vector $X^l = \left[x_1^l, x_2^l, \ldots, x_K^l\right]^T$, whose dimension K corresponding to the number of frequencies used, represents the complex normalized scattering coefficient of a prototype from an unknown class. The complex number x_k^l corresponds to the complex scattering coefficient whose magnitude is the square root of the measured cross section in square meters, m^2, and whose phase is that of the measured signal.

For the simulation experiments of both parametric and nonparametric sequential techniques, it is assumed that the observation process corresponds to a linear system measurement of the signal vector, S, in the presence of additive Gaussian noise. The signal vector is taken as one of the N_s prototypes from one of the M classes, where each class corresponds to one of the five aircraft in the database.

The additive Gaussian noise is represented by two uncorrelated random numbers W_R, and W_I, each having a Gaussian distribution with zero mean and variance $\frac{\sigma^2}{2}$. Thus the total additive noise $W = W_R + jW_I$ ($j = \sqrt{-1}$) has a Gaussian distribution with zero mean and variance σ^2. The l^{th} observed complex normalized scattering coefficient of a target of class i and at j^{th} azimuth position and using k^{th} frequency component is:

$$x_k^l = \left[Re(s_{i,j,k}) + W_R^l\right] + j\left[Im(s_{i,j,k}) + W_I^l\right] \tag{5.1}$$

Table 1: Error rate and average number of measurements for the Reed technique with $M = 5$ classes, $N_s = 6$ prototypes/class, and $K = 4$.

Noise (dBsm)	Error	$E\{n\}$
16	0.000	2.14
19	0.004	2.44
22	0.044	3.29
25	0.136	4.57
28	0.288	6.86
31	0.412	8.90

5.3 Simulation Approach

When the azimuth angle of the radar object is assumed to be known, then $N_s = 1$ and the resulting vector observation, $X = S+W$ is assumed to be Gaussian random vector with mean vector S and covariance matrix, $\sigma^2 I$, where I is the $K \times K$ identity matrix. When the azimuth angle of the object is assumed to be unknown, or known to within a specified range, the observation vector X is assumed to be distributed as a Gaussian mixture as in (2.4), corresponding to the N_s subclasses for each of the M classes. The target is assumed to stay in the same azimuth position whenever a new observation is requested, that is, all measurements correspond to one target at a fixed position in azimuth and elevation.

In Figure 6, the probability of classification error is shown as a function of noise power level for the Reed test using the modified form of the thresholds (3.7). The average number of required measurements, $E\{n\}$ for this test as a function of noise power level is given in Figure 7. These results are also tabulated in Table 1.

The results in Figures 6-7 are computed for $N_s = 6$ prototypes per class, corresponding to target azimuth angles of 0°, 10°, 20°, 30°, 40°, and 50°. In this

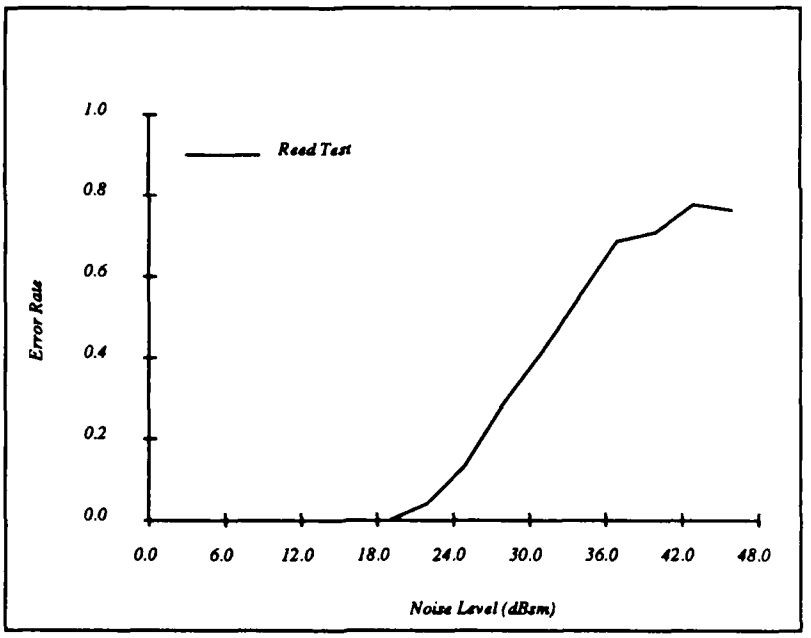


Figure 6: Error rates for the Reed test with $M = 5$ classes, $N_s = 6$ prototypes/class, $K = 4$.

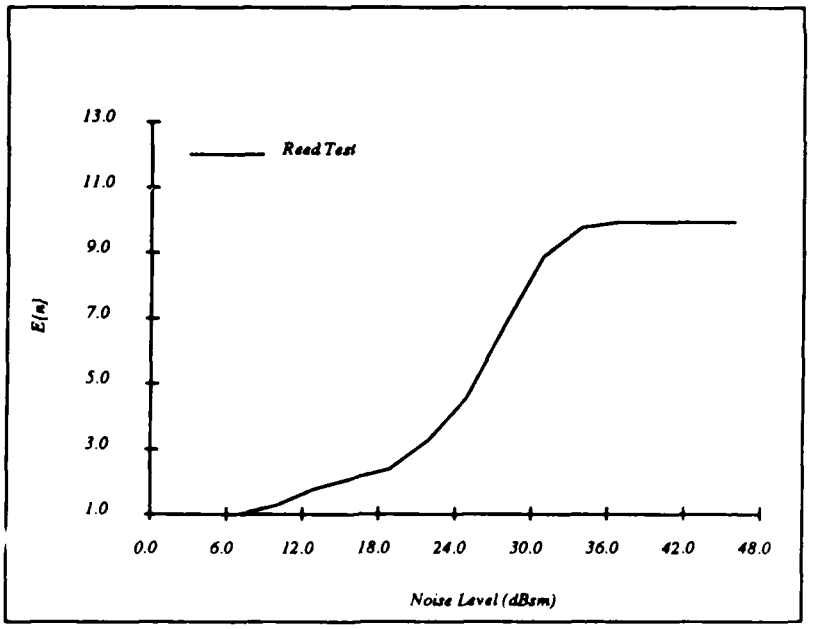


Figure 7: Average number of measurements for the Reed test with $M = 5$ classes, $N_s = 6$ prototypes/class, $K = 4$.

case, the measurement vectors are of dimension, $K = 4$ corresponding to coherent backscatter measurements at 8, 11, 16, and 25 MHz. The error percentages and the average number of measurements are based on the results of three hundred experiments.

The coordinates of the abscissa in these figures refer to the power, or variance, of the Gaussian observation in terms of decibels relative to the power, P_r received from an ideal radar signal reflector with one square meter area, i.e.,

$$\text{Noise Power}(\text{dB}m^2) = 10 \cdot \log_{10}\left(\frac{\sigma^2}{P_r}\right) \tag{5.2}$$

For the experimental results presented here, the average signal power of the component of the radar measurement due to the target of interest is approximately $20\text{dB}m^2$, so that an approximate SNR in decibels may be calculated for any of the data presented below as $SNR \approx 20$ - Noise Power($\text{dB}m^2$). Figures 8-9 show the effect of the suggested modifications to the Reed test on both the probability of error and average number of measurements, with $N_s = 6$, $K = 4$.

In Figures 10-12 the probability of classification error is shown as a function noise power level for the Reed test with $N_s = 5(4)$, and $K = 3(2)$ respectively. The average number of required measurements, for these cases are shown in Figures 11-13. Figures 14-15 show the classification error and average number of measurements for the Reed test with $N_s = 2$ $(0°, 10°)$ and $K = 2$.

Figures 16-17 show the probability of error and the average number of measurements as a function of the noise power level for the Reed test using a single frequency radar with $N_s = 1$ (simple classes). Similar results for the Armitage and Palmer tests are shown in Figures 18-37. These results are also tabulated in Tables 2- 3.

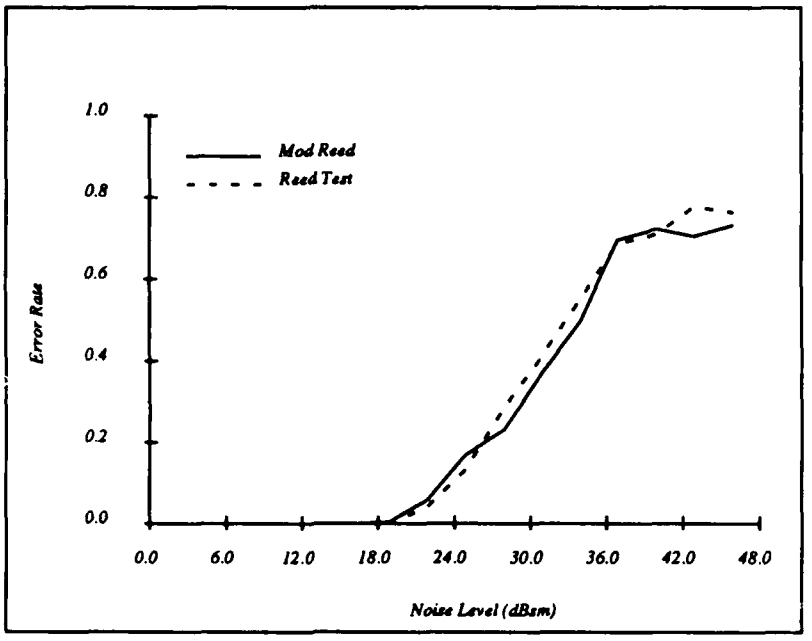


Figure 8: Comparison of error rate for Reed and modified Reed techniques with $M = 5$ classes, $N_s = 6$ prototypes/class, $K = 4$.

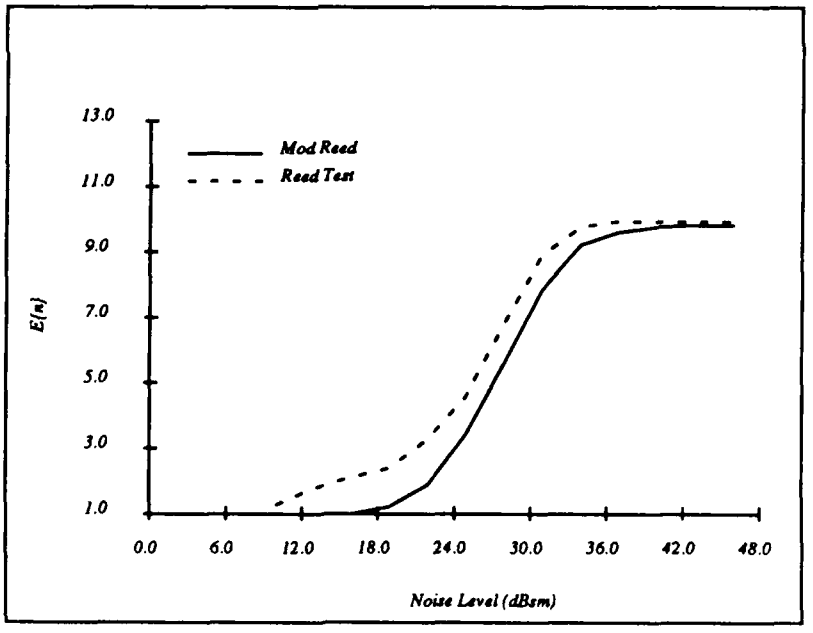


Figure 9: Comparison of the average number of measurements for the Reed test and the modified Reed with $M = 5$ classes, $N_s = 6$ prototypes/class, $K = 4$.

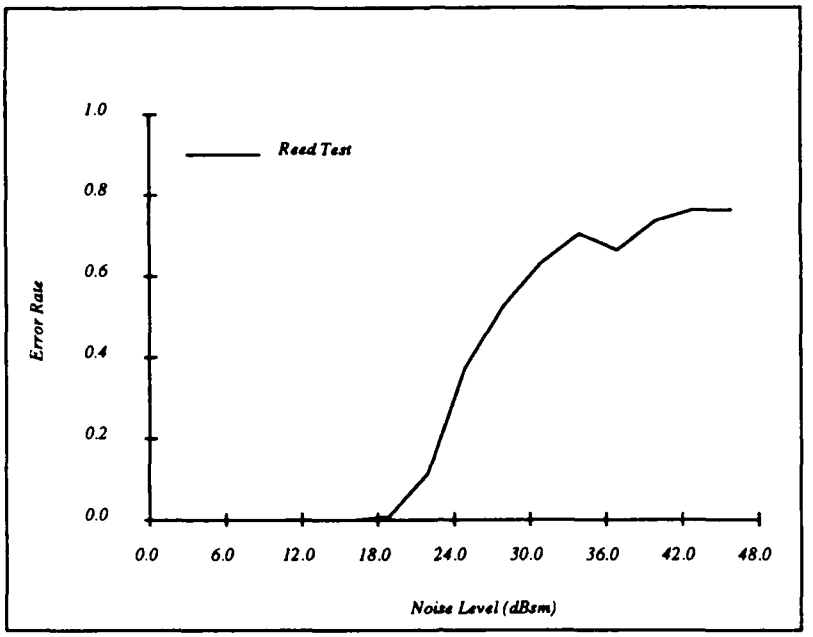


Figure 10: Error rates for the Reed test with $M = 5$ classes, $N_s = 5$ prototypes/class, $K = 3$.

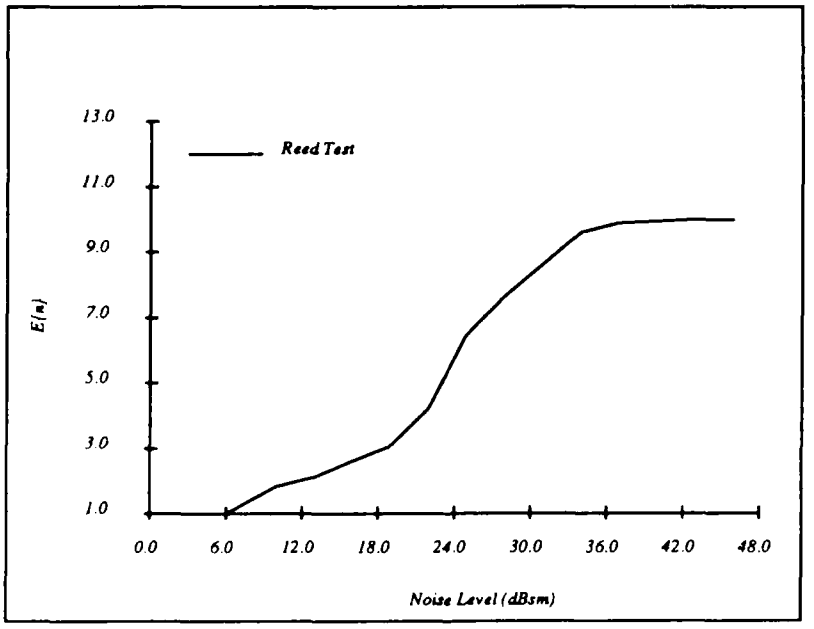


Figure 11: Average number of measurements for the Reed test with $M = 5$ classes, $N_s = 5$ prototypes/class, $K = 3$.

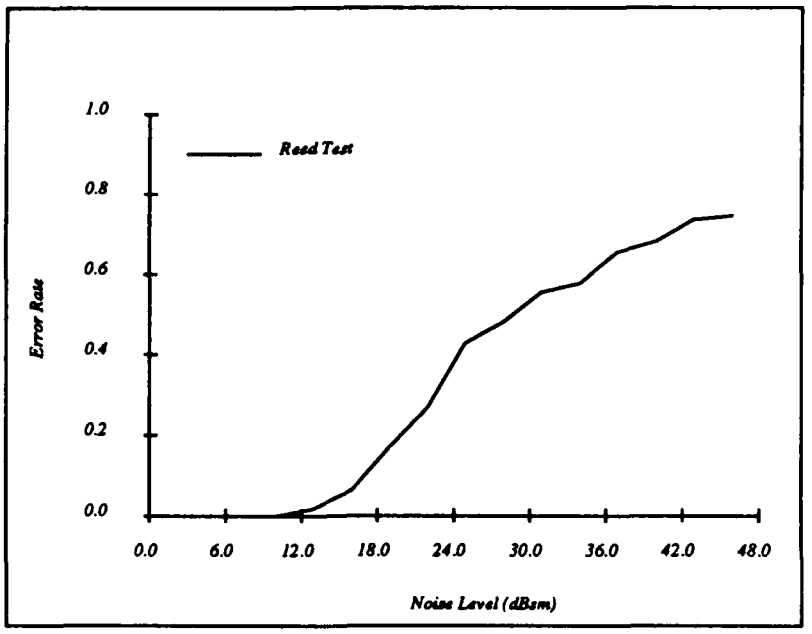


Figure 12: Error rates for the Reed test with $M = 5$ classes, $N_s = 4$ prototypes/class, $K = 2$.

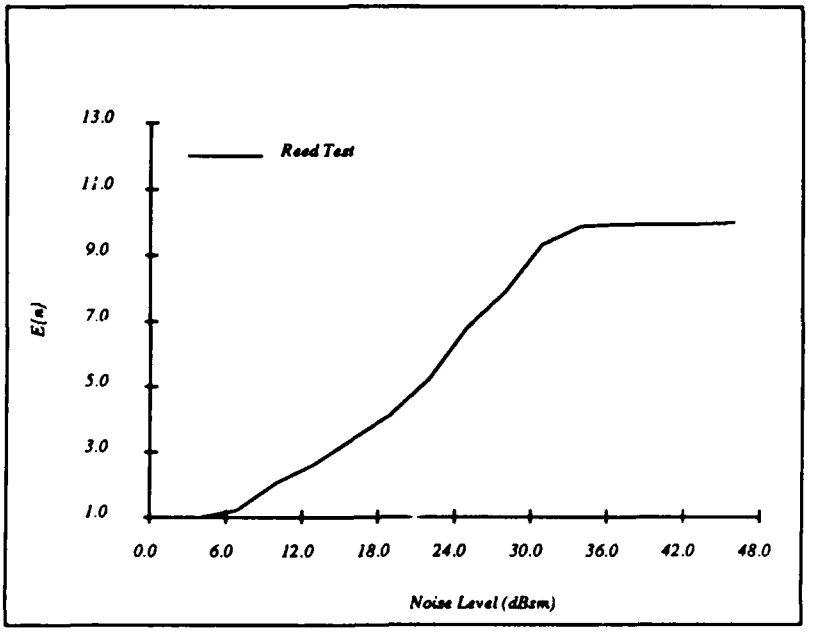


Figure 13: Average number of measurements for the Reed test with $M = 5$ classes, $N_s = 4$ prototypes/class, $K = 2$.

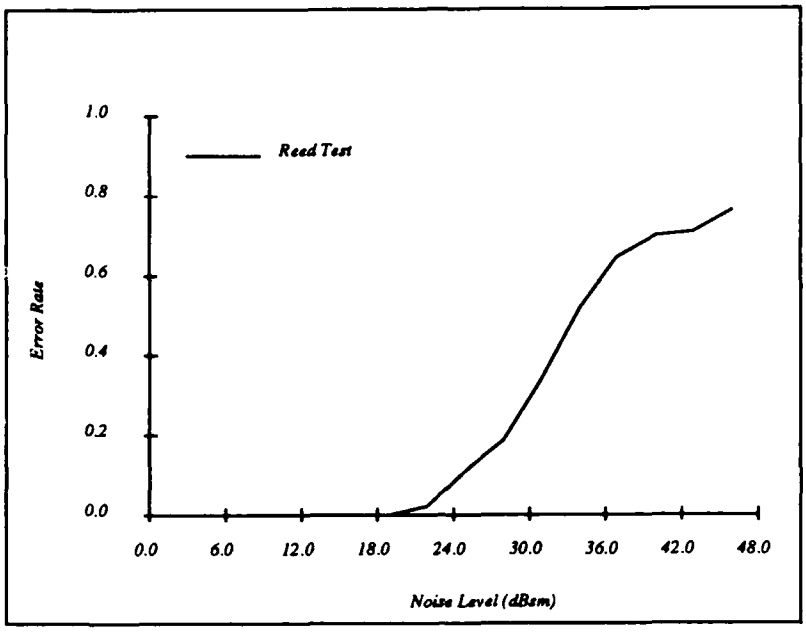


Figure 14: Error rates for the Reed test with $M = 5$ classes, $N_s = 2$ prototypes/class, $K = 2$.

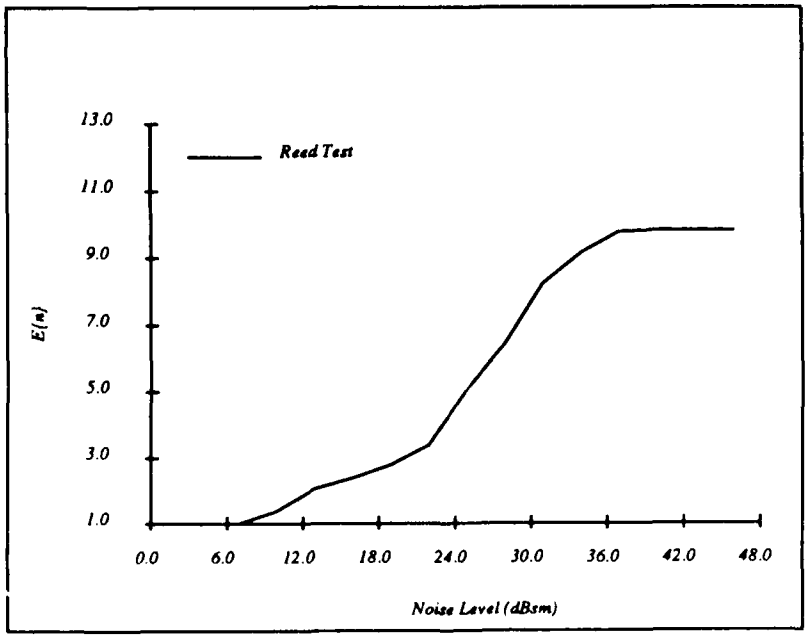


Figure 15: Average number of measurements for the Reed test with $M = 5$ classes, $N_s = 2$ prototypes/class, $K = 2$.

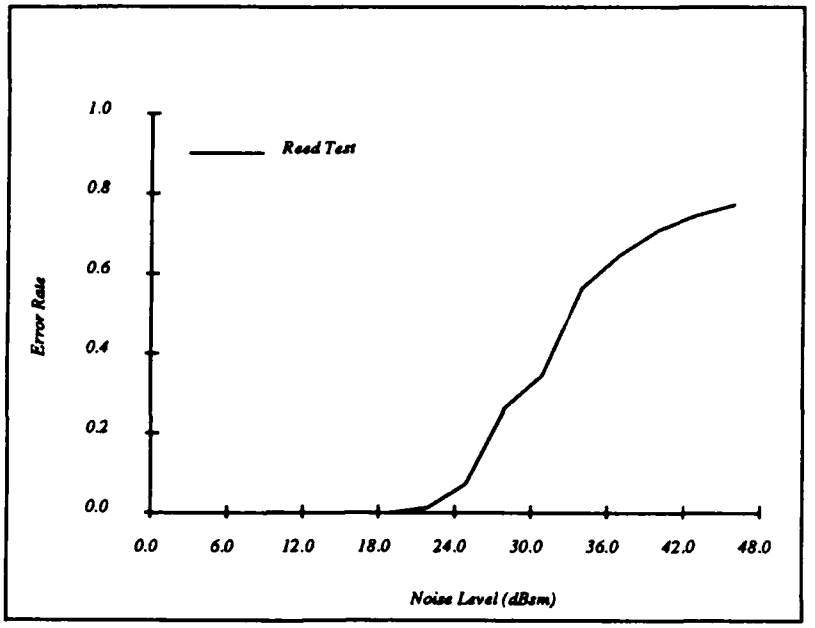


Figure 16: Error rates for the Reed test with $M = 5$ classes, $N_s = 1$ prototypes/class, $K = 1$.

Table 2: Error rate and average number of measurements for the Armitage technique with $M = 5$ classes, $N_s = 6$ prototypes/class, and $K = 4$.

Noise (dBsm)	Error	$E\{n\}$
16	0.000	1.37
19	0.000	2.28
22	0.006	4.50
25	0.064	7.18
28	0.234	9.41
31	0.352	9.96

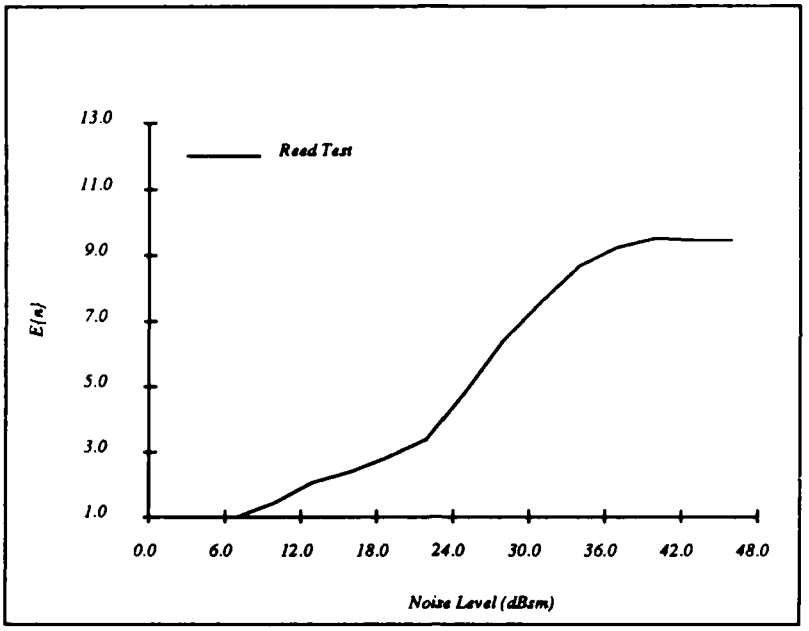


Figure 17: Average number of measurements for the Reed test with $M = 5$ classes, $N_s = 1$ prototypes/class, $K = 1$.

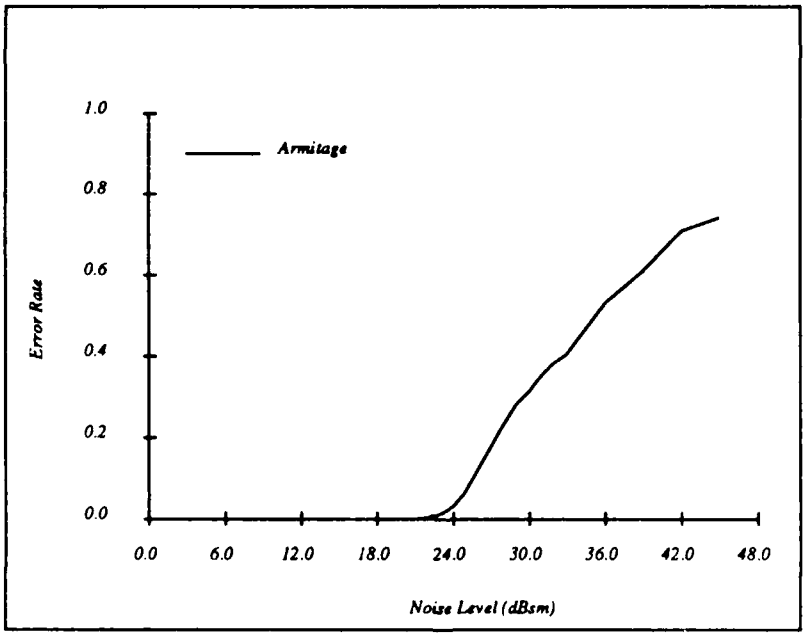


Figure 18: Error rates for the Armitage test with $M = 5$ classes, $N_s = 6$ prototypes/class, $K = 4$.

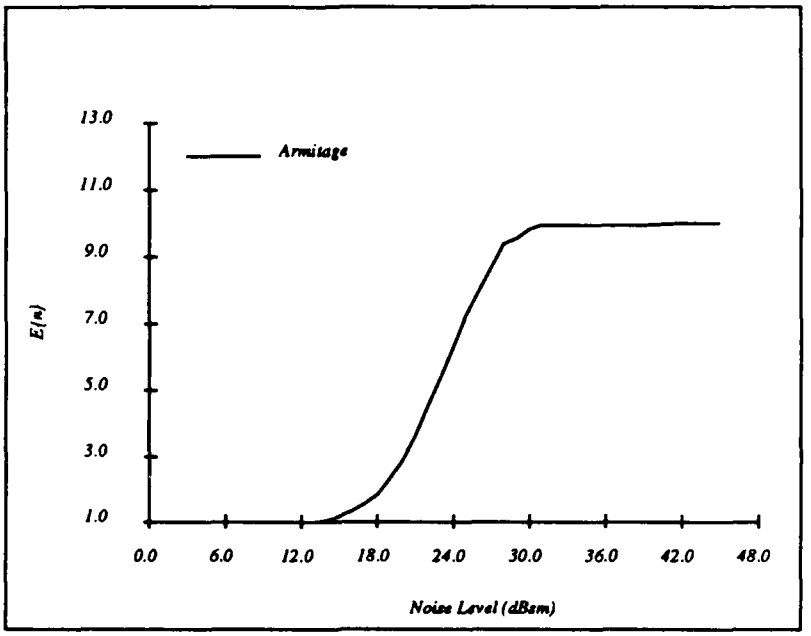


Figure 19: Average number of measurements for the Armitage test with $M = 5$ classes, $N_s = 6$ prototypes/class, $K = 4$.

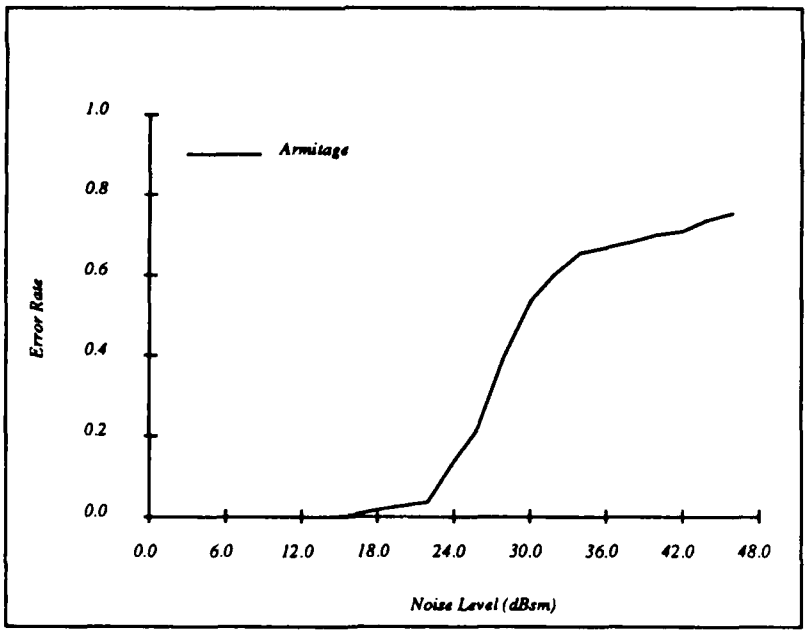


Figure 20: Error rates for the Armitage test with $M = 5$ classes, $N_s = 5$ prototypes/class, $K = 3$.

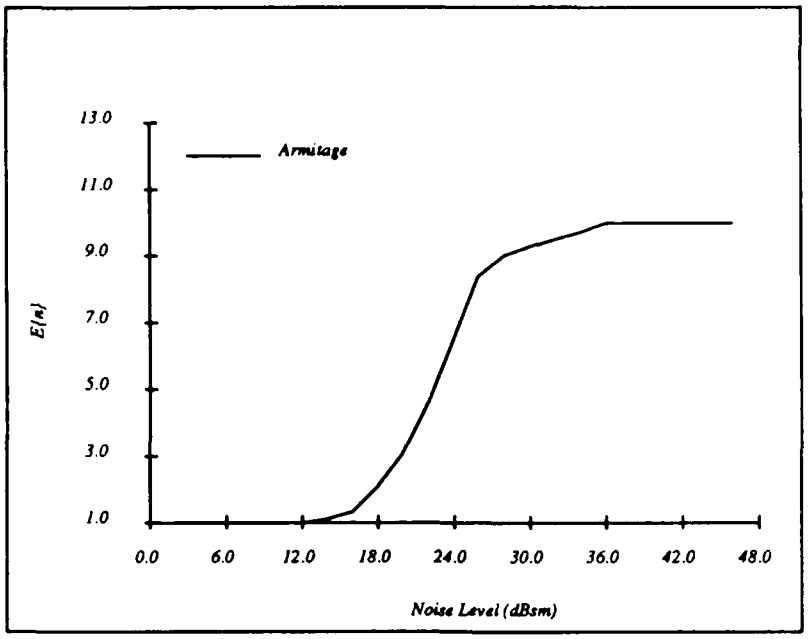


Figure 21: Average number of measurements for the Armitage test with $M = 5$ classes, $N_s = 5$ prototypes/class, $K = 3$.

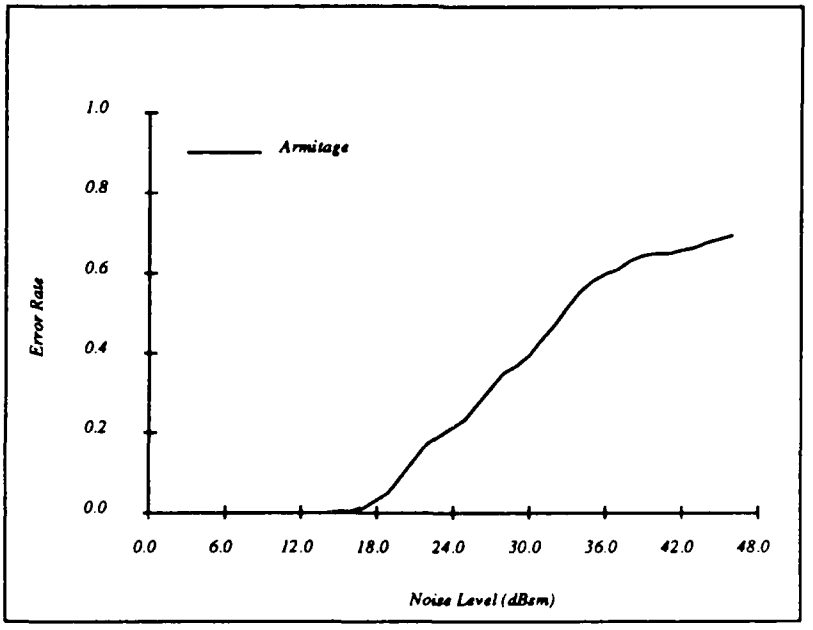


Figure 22: Error rates for the Armitage test with $M = 5$ classes, $N_s = 4$ prototypes/class, $K = 2$.

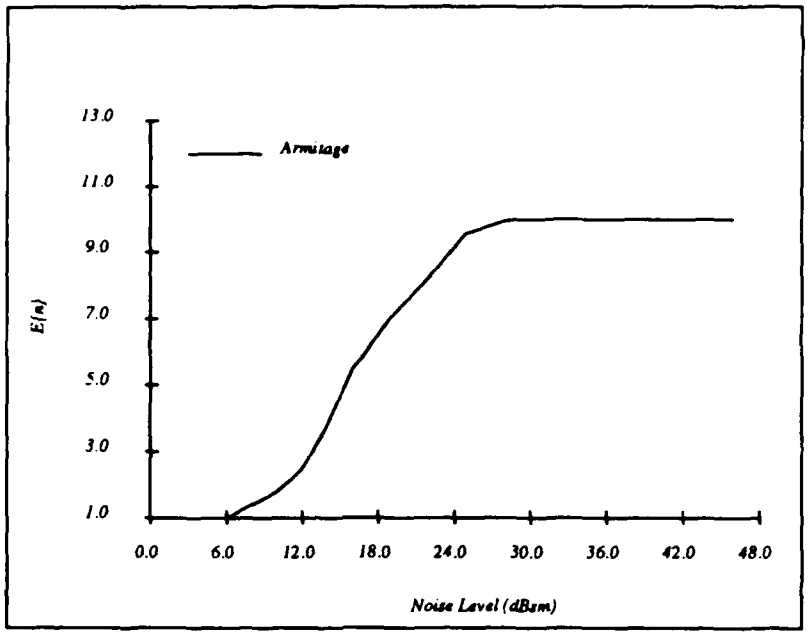


Figure 23: Average number of measurements for the Armitage test with $M = 5$ classes, $N_s = 4$ prototypes/class, $K = 2$.

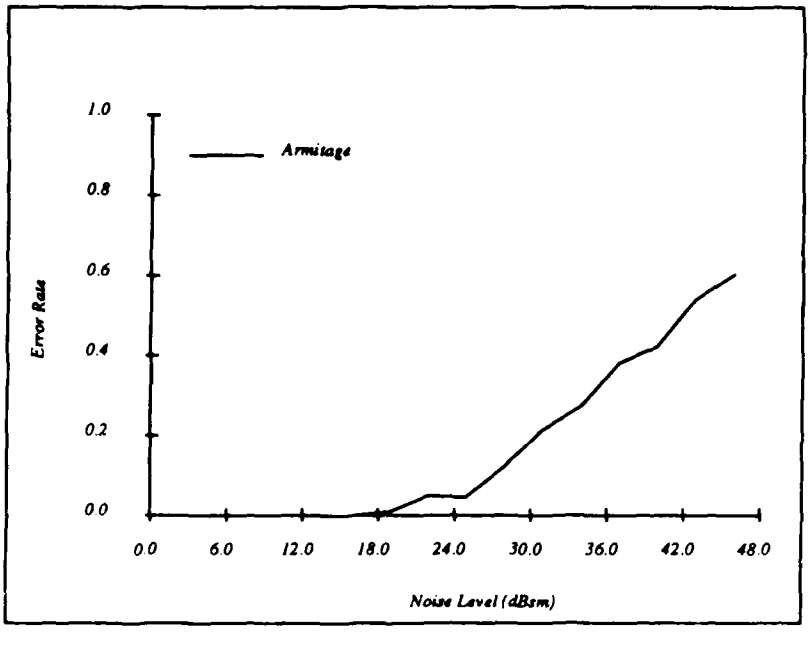


Figure 24: Error rates for the Armitage test with $M = 5$ classes, $N_s = 1$ prototypes/class, $K = 1$.

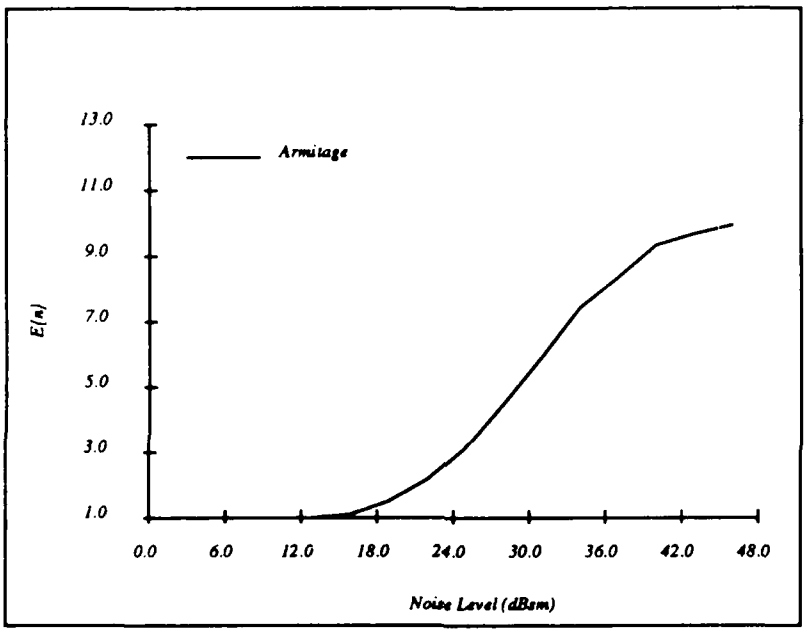


Figure 25: Average number of measurements for the Armitage test with $M = 5$ classes, $N_s = 1$ prototypes/class, $K = 1$.

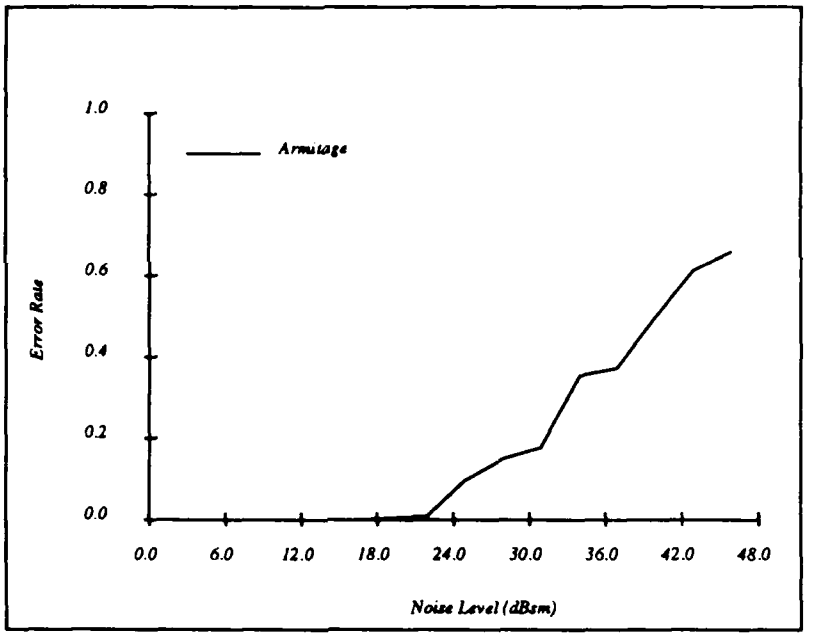


Figure 26: Error rates for the Armitage test with $M = 5$ classes, $N_s = 2$ prototypes/class, $K = 2$.

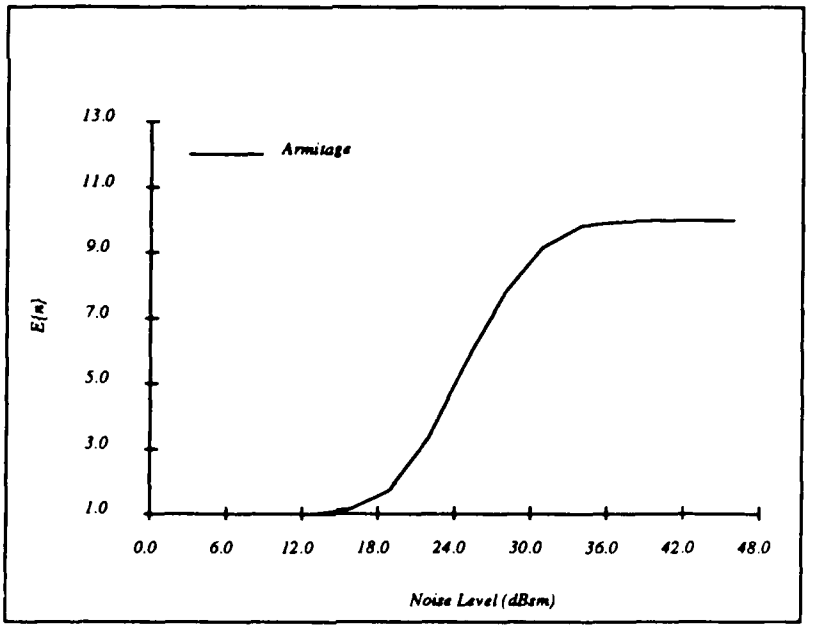


Figure 27: Average number of measurements for the Armitage test with $M = 5$ classes, $N_s = 2$ prototypes/class, $K = 2$.

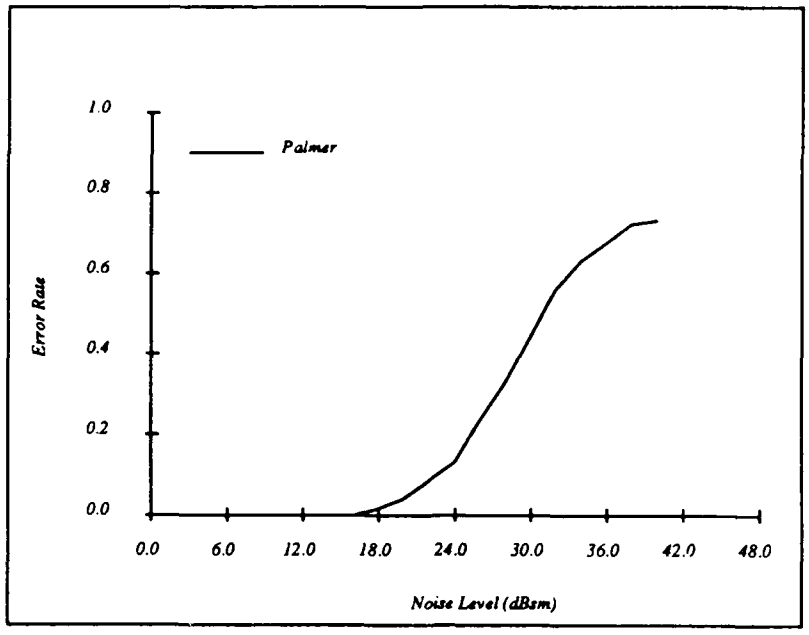


Figure 28: Error rates for the Palmer test with $M = 5$ classes, $N_s = 6$ prototypes/class, $K = 4$.

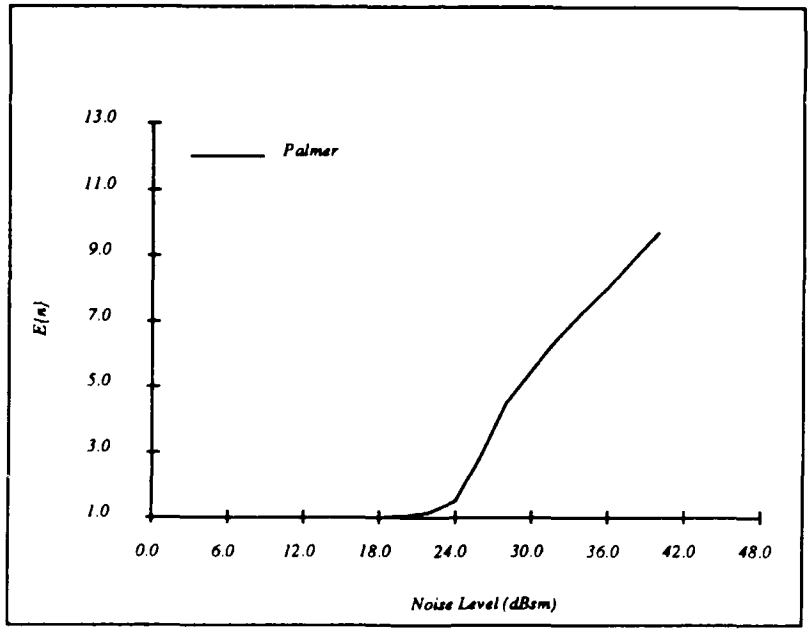


Figure 29: Average number of measurements for the Palmer test with $M = 5$ classes, $N_s = 6$ prototypes/class, $K = 4$.

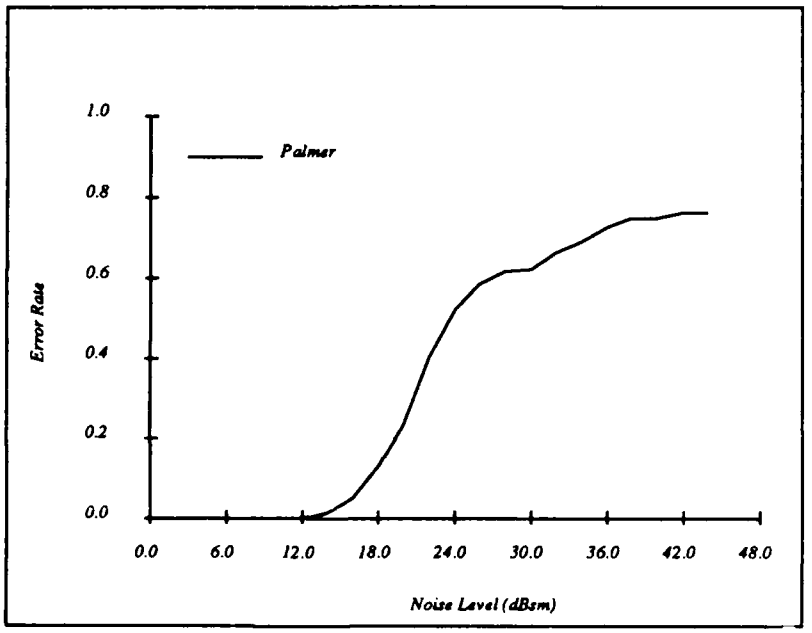


Figure 30: Error rates for the Palmer test with $M = 5$ classes, $N_s = 5$ prototypes/class, $K = 3$.

Figures 38-39, show the effect of the proposed modification (3.4) to the Armitage procedure. For these simulations, target prototypes at azimuth angles of 0°, 10°, 20°, 30°, and 40° are used for measurement vectors of dimension $K = 3$ frequencies at 8, 9, and 10 MHz. These results are also tabulated in Tables 4-5-6. From these figures it is clear that fewer measurements are required for the modified test for higher values of the parameter r while the resulting error percentages remain almost the same as for the original test ($r = 0$).

Figures 40-41 show the reduction in the average number of obsevations due to the modified armitage thresholds (3.4) where the target prototypes are at azimuth angles of 0°, 10°, 20°, 30°. The measurement vector is of dimension $K = 2$ frequencies at 8 and 9 MHz.

In Figure 42, the probability of classification error is shown as a function of noise power level for the tree structured sequential test. The average number of required measurements, $E\{n\}$ for this test as a function of noise power level is given in Figure 43. This sequential test starts by classifying the target into groups of hypotheses $\{\omega_5\}$, or $\{\omega_1, \omega_2, \omega_3, \omega_4\}$ then, if it is not of class $\{\omega_5\}$, the target is classified as a member of the groups $\{\omega_1, \omega_2\}$ or $\{\omega_3, \omega_4\}$. Finally, the classification

Table 3: Error rate and average number of measurements for the Palmer technique with $M = 5$ classes, $N_s = 6$ prototypes/class, and $K = 4$.

Noise (dBsm)	Error	$E\{n\}$
16	0.003	1.01
18	0.020	1.03
22	0.088	1.20
26	0.235	2.84
28	0.329	4.51
30	0.448	5.52

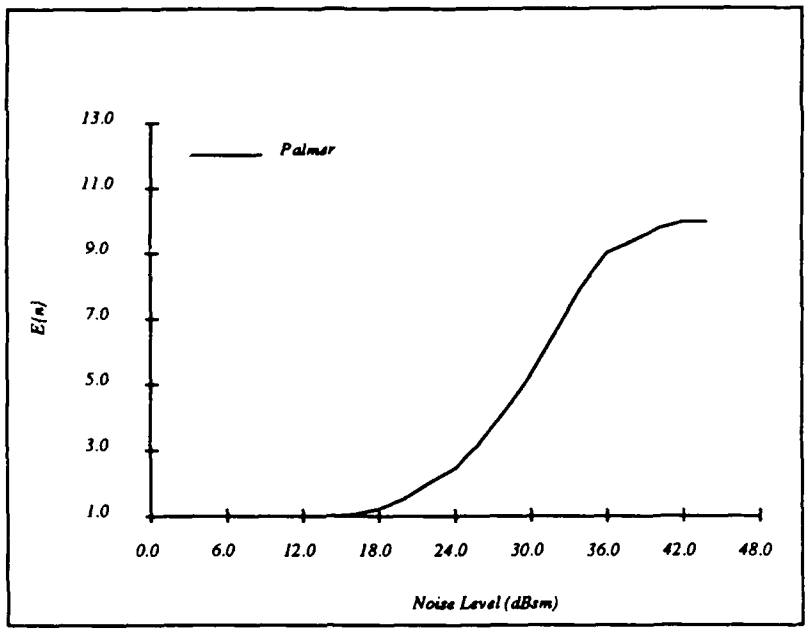


Figure 31: Average number of measurements for the Palmer test with $M = 5$ classes, $N_s = 5$ prototypes/class, $K = 3$.

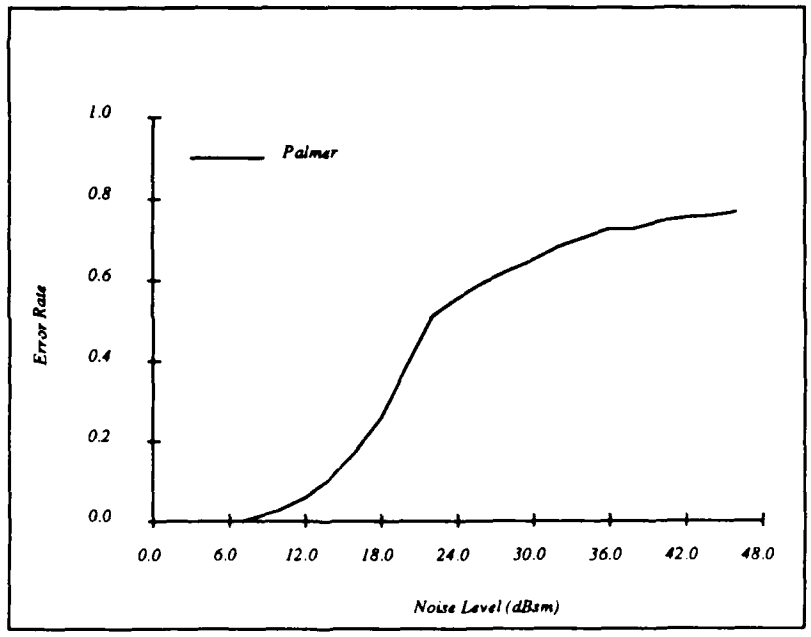


Figure 32: Error rates for the Palmer test with $M = 5$ classes, $N_s = 4$ prototypes/class, $K = 2$.

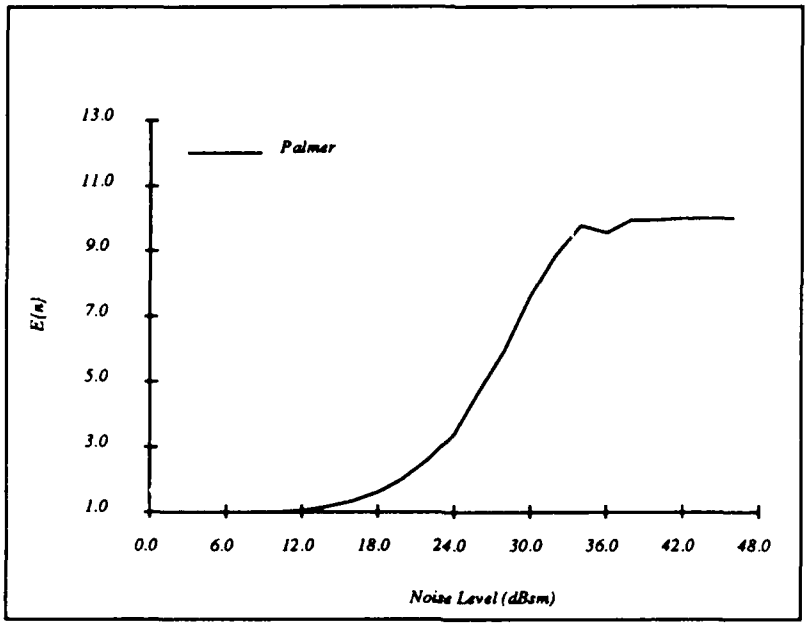


Figure 33: Average number of measurements for the Palmer test with $M = 5$ classes, $N_s = 4$ prototypes/class, $K = 2$.

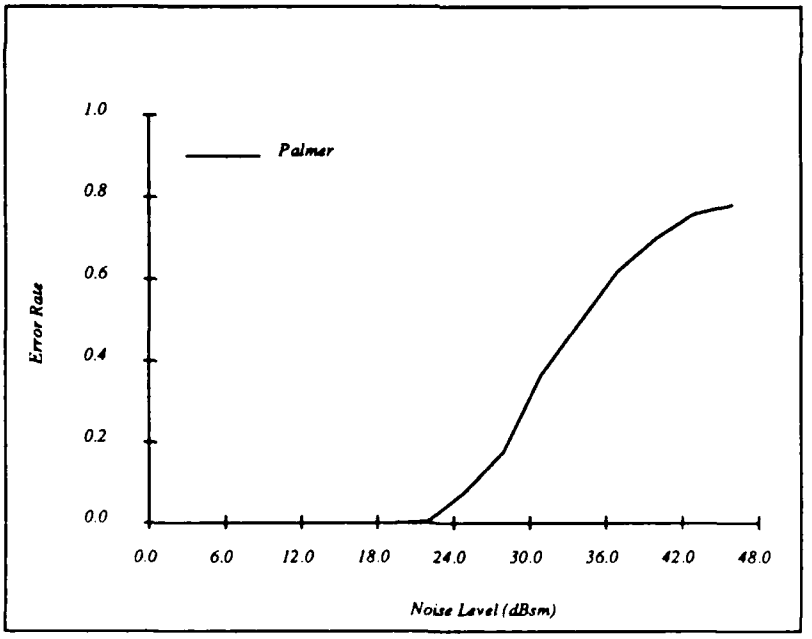


Figure 34: Error rates for the Palmer test with $M = 5$ classes, $N_s = 1$ prototypes/class, $K = 1$.

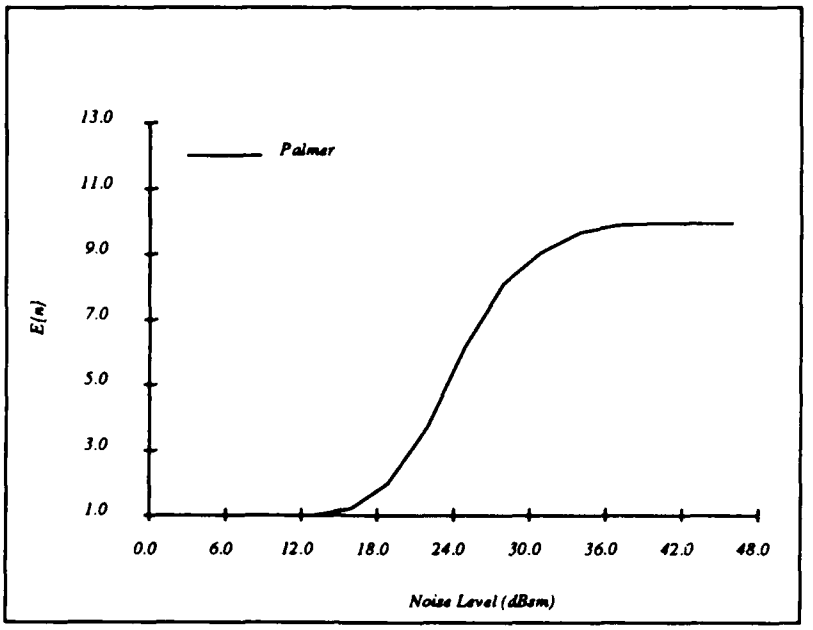


Figure 35: Average number of measurements for the Palmer test with $M = 5$ classes, $N_s = 1$ prototypes/class, $K = 1$.

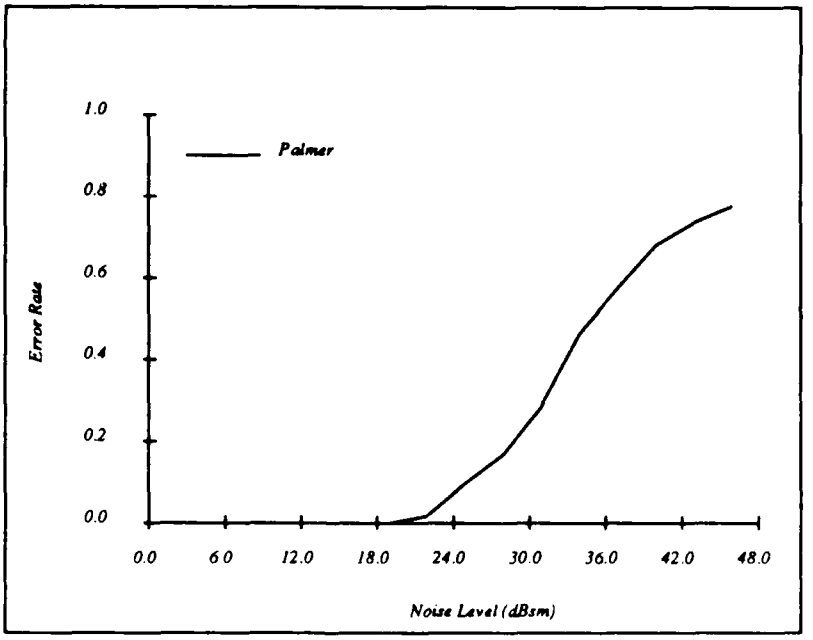


Figure 36: Error rates for the Palmer test with $M = 5$ classes, $N_s = 2$ prototypes/class, $K = 2$.

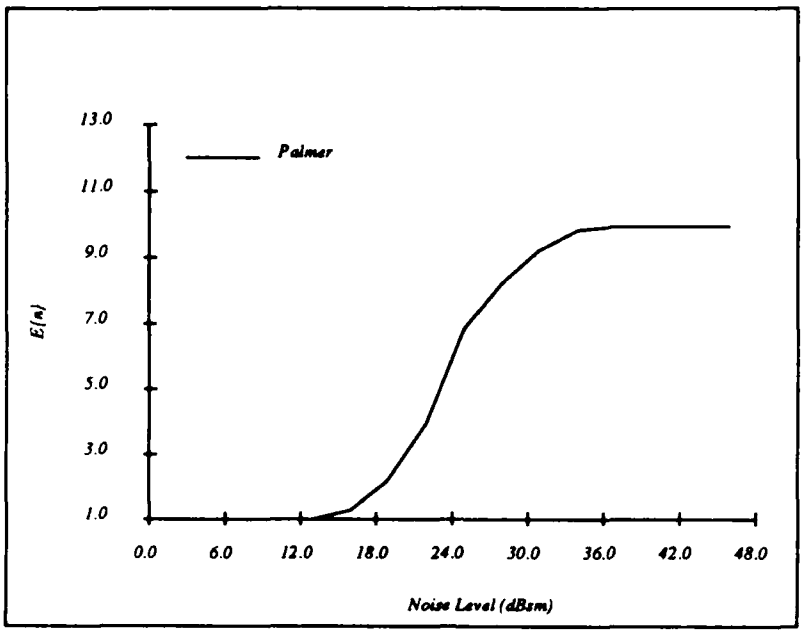


Figure 37: Average number of measurements for the Palmer test with $M = 5$ classes, $N_s = 2$ prototypes/class, $K = 2$.

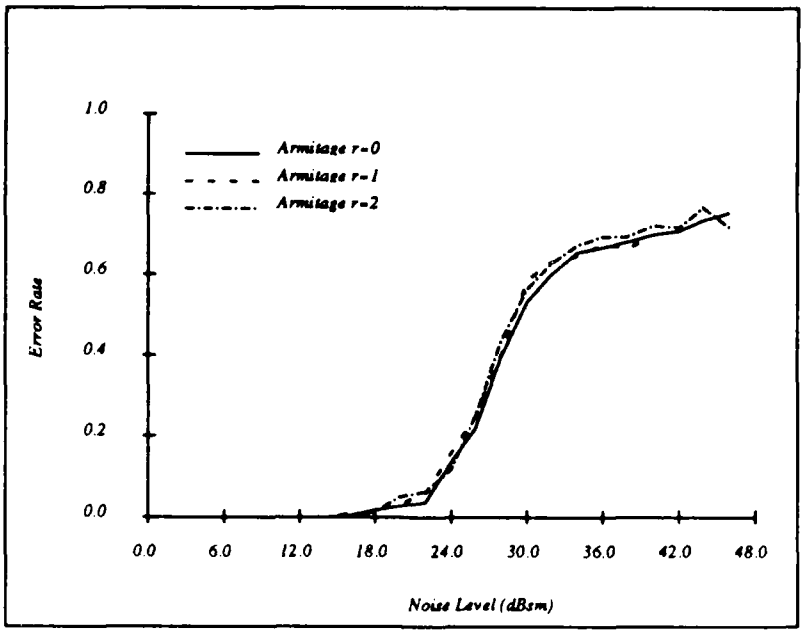


Figure 38: Error rates for the modified Armitage thresholds for $r = 0, 1, 2$ with $M = 5$ classes, $N_s = 5$ prototypes/class, $K = 3$.

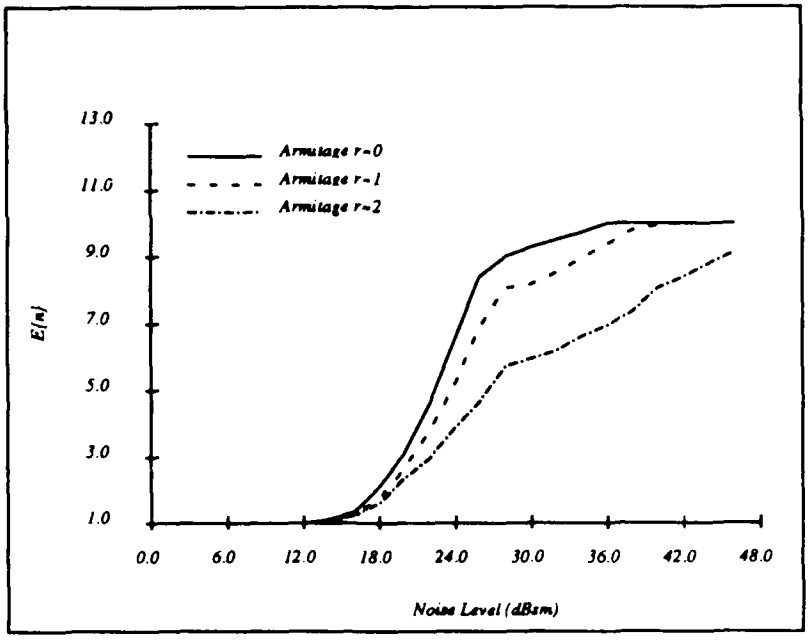


Figure 39: Average number of measurements for the modified Armitage thresholds for $r = 0, 1, 2$ with $M = 5$ classes, $N_s = 5$ prototypes/class, $K = 3$.

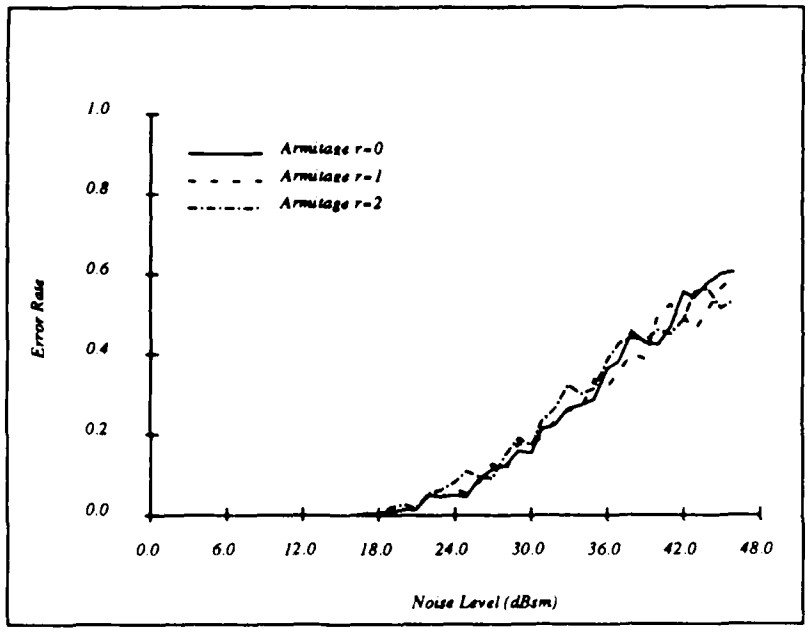


Figure 40: Error rates for the modified Armitage thresholds for $r = 0, 1, 2$ with $M = 5$ classes, $N_s = 1$ prototypes/class, $K = 1$.

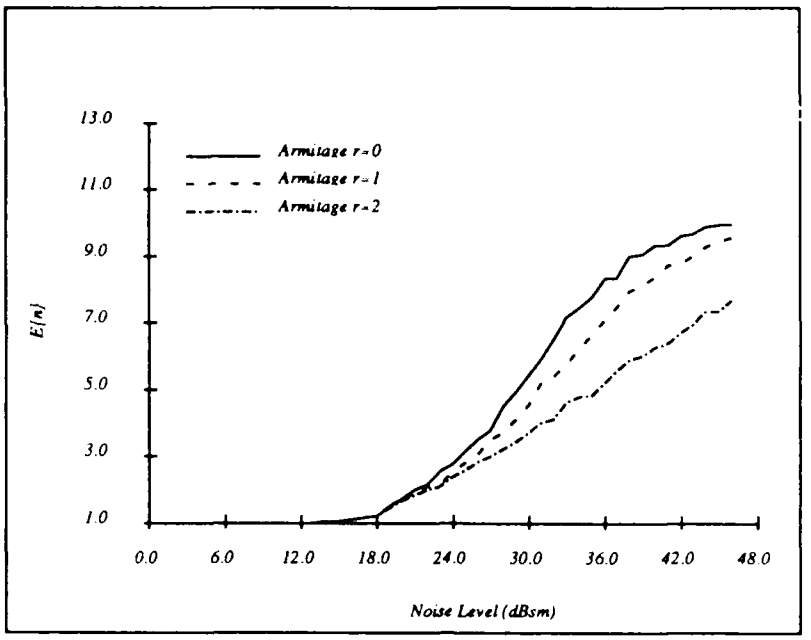


Figure 41: Average number of measurements for the modified Armitage thresholds for $r = 0, 1, 2$ with $M = 5$ classes, $N_s = 1$ prototypes/class, $K = 1$.

Table 4: Error rate and average number of measurements for the modified Armitage thresholds with $M = 5$ classes, $N_s = 5$ prototypes/class, $K = 3$ and $r = 0$.

Noise (dBsm)	Error	$E\{n\}$
30	0.536	9.292
32	0.600	9.552
34	0.656	9.760
36	0.672	10.00
38	0.684	10.00
40	0.704	10.00

Table 5: Error rate and average number of measurements for the modified Armitage thresholds with $M = 5$ classes, $N_s = 5$ prototypes/class, $K = 3$ and $r = 1$.

Noise (dBsm)	Error	$E\{n\}$
30	0.588	8.180
32	0.636	8.548
34	0.648	8.996
36	0.680	9.404
38	0.664	9.808
40	0.704	9.992

Table 6: Error rate and average number of measurements for the modified Armitage thresholds with $M = 5$ classes, $N_s = 5$ prototypes/class, $K = 3$ and $r = 2$.

Noise (dBsm)	Error	$E\{n\}$
30	0.568	5.964
32	0.628	6.212
34	0.676	6.652
36	0.696	6.948
38	0.696	7.364
40	0.724	8.052

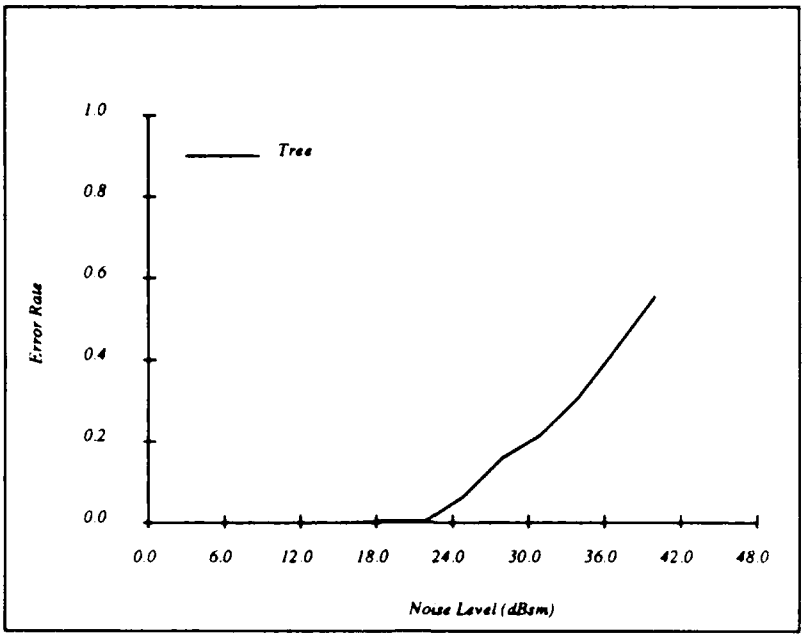


Figure 42: Error rates for the tree structured sequential test test with $M = 5$ classes, $N_s = 6$ prototypes/class, $K = 4$.

is to any of the two classes in either of these groups. The thresholds used in this simulation are of the form shown in Figure 3. That is, truncation takes place once and the test continues after that as nonsequential binary likelihood ratio tests. Tabulated results are shown in Table 7.

In Figure 44, the probability of classification error is shown as a function of noise power level for the sequential nearest neighbor test using the modified form of the thresholds (4.3). The average number of required measurements, $E\{n\}$ for this test as a function of noise power level is given in Figure 45. These results are also tabulated in Table 8.

The results in this figure are computed for $N_s = 6$ prototypes per class, corresponding to target azimuth angles of 0°, 10°, 20°, 30°, 40°, and 50°. In this case, the measurement vectors are of dimension, $K = 4$ corresponding to coherent

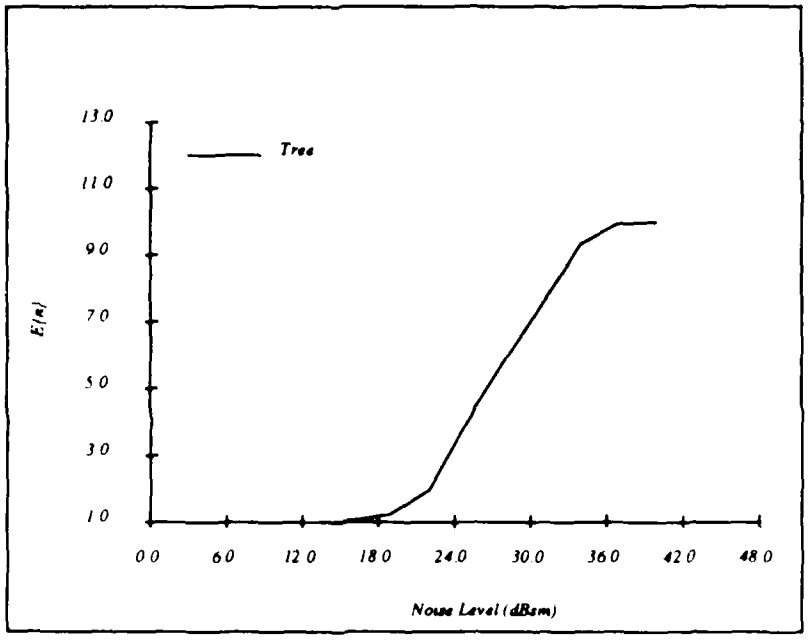


Figure 43: Average number of measurements for the tree structured sequential test with $M = 5$ classes, $N_s = 6$ prototypes/class, $K = 4$.

Table 7: Error rate and average number of measurements for the tree structured sequential technique with $M = 5$ classes, $N_s = 6$ prototypes/class, and $K = 4$.

Noise (dBsm)	Error	$E\{n\}$
16	0.002	1.05
19	0.008	1.27
22	0.008	1.99
25	0.068	3.95
28	0.161	5.84
31	0.216	7.51

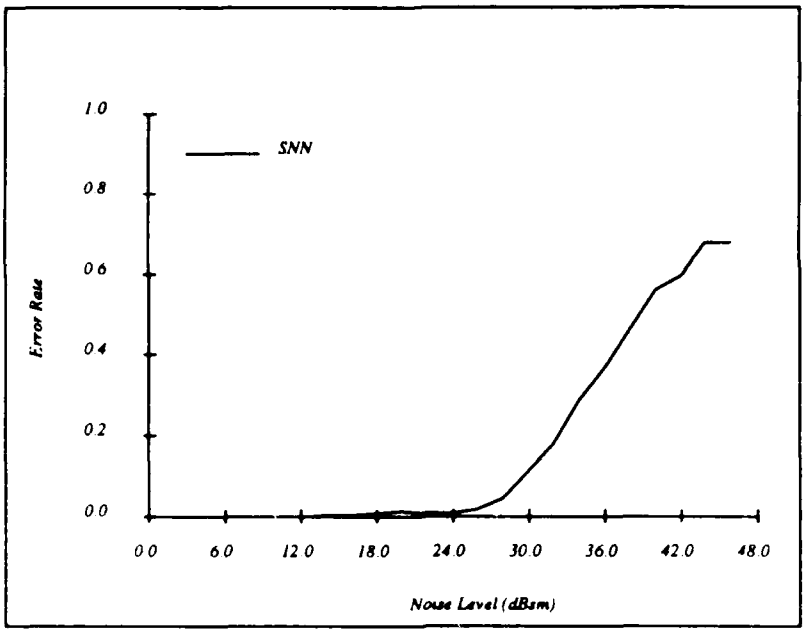


Figure 44: Error rates for the sequential nearest neighbor test with $M = 5$ classes, $N_s = 6$ prototypes/class, $K = 4$.

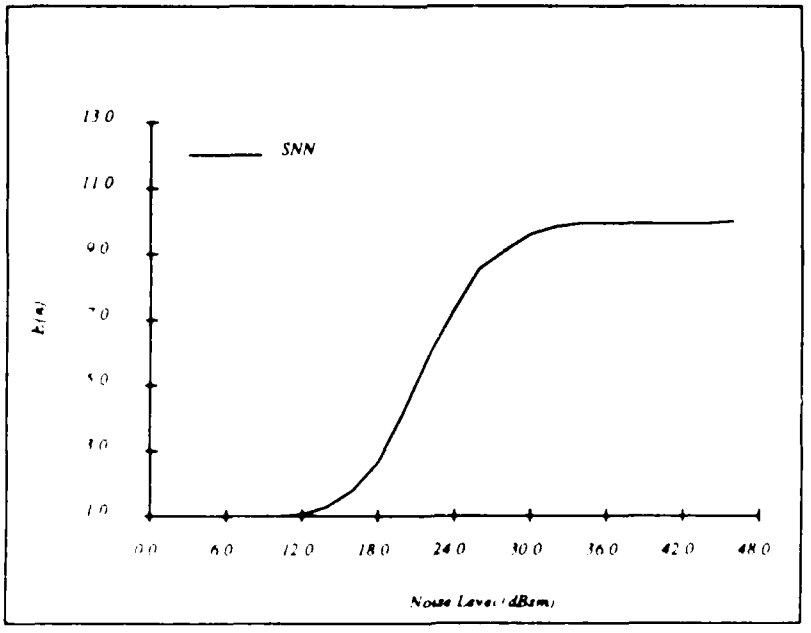


Figure 45: Average number of measurements for the sequential nearest neighbor test with $M = 5$ classes, $N_s = 6$ prototypes/class, $K = 4$.

Table 8: Error rate and average number of measurements for the sequential nearest neighbor technique with $M = 5$ classes, $N_s = 6$ prototypes/class, and $K = 4$.

Noise (dBsm)	Error	$E\{n\}$
16	0.007	1.78
18	0.010	2.71
20	0.014	4.13
22	0.012	5.88
24	0.012	7.38
26	0.020	8.56

backscatter measurements at 8, 11, 16, and 25 MHz. The error percentages and the average number of measurements are based on the results of three hundred experiments.

In Figures 46-48, the probability of classification error is shown as a function noise power level for the sequential nearest neighbor test with $N_s = 5(4)$, and $K = 3(2)$ respectively. The average number of required measurements, for these cases are shown in Figures 47-49. Figures 50-51 show the probability of error and the average number of measurements as a function of the noise power level for the sequential nearest neighbor test using a single frequency radar with $N_s = 1$ (simple classes). Figures 52-53 show the misclassification error and the average number of observations for the sequential nearest neighbor with $N_s = 2$ $(0°, 10°)$ prototypes and $K = 2$.

Figures 56-57, show the probability of error and the average number of observations as a function of the threshold A_i (accepted minimum distance) for the sequential nearest neighbor test. These experiments were run at a 25 (dBsm) noise power level.

A comparison between the sequential nearest neighbor and the nearest neigh

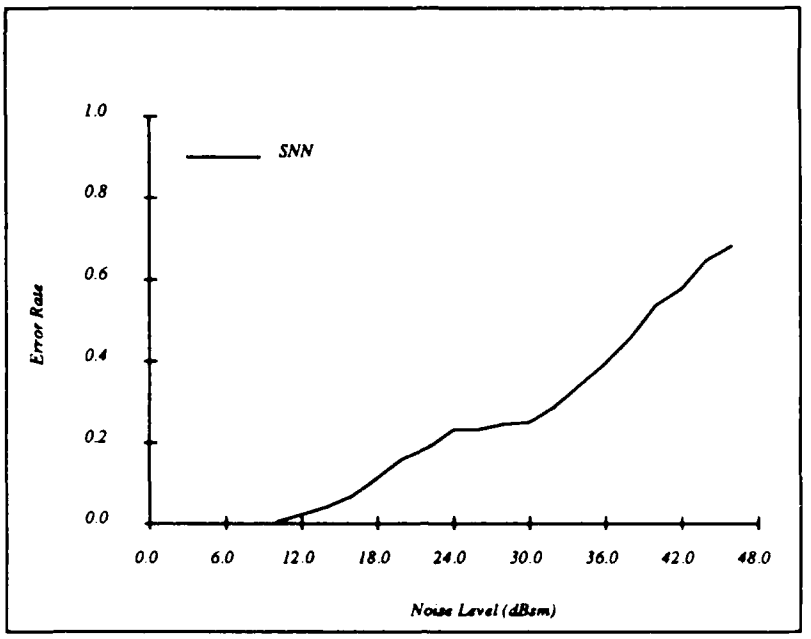


Figure 46: Error rates for the sequential nearest neighbor test with $M = 5$ classes, $N_s = 5$ prototypes/class, $K = 3$.

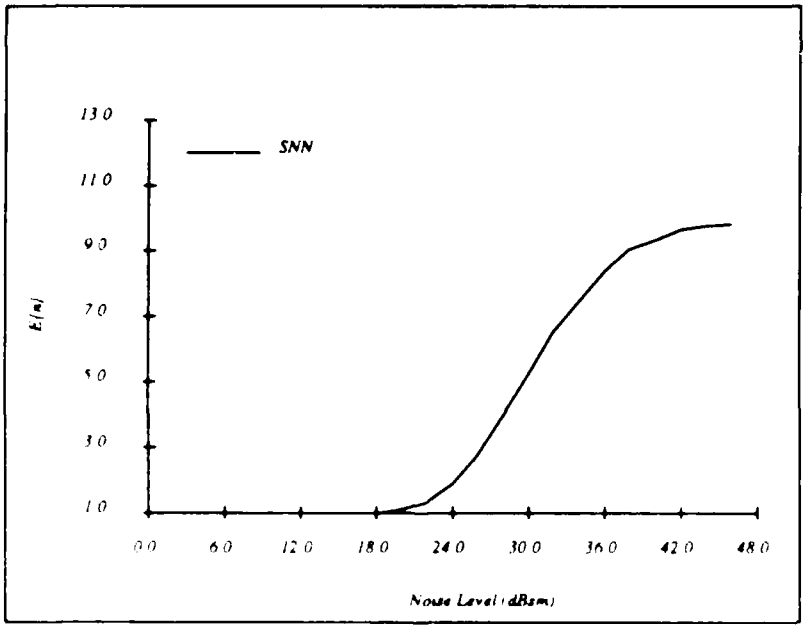


Figure 47: Average number of measurements for the sequential nearest neighbor test with $M = 5$ classes, $N_s = 5$ prototypes/class, $K = 3$.

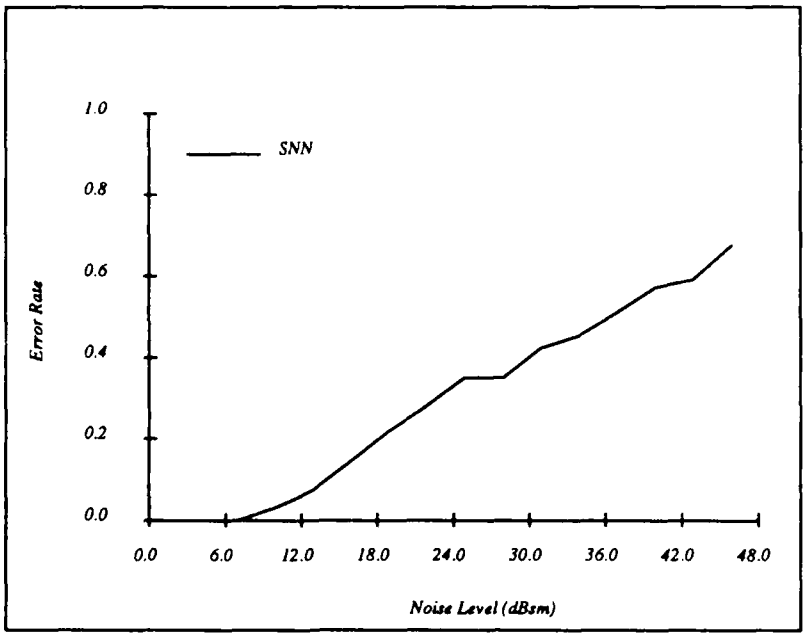


Figure 48: Error rates for the sequential nearest neighbor test with $M = 5$ classes, $N_s = 4$ prototypes/class, $K = 2$.

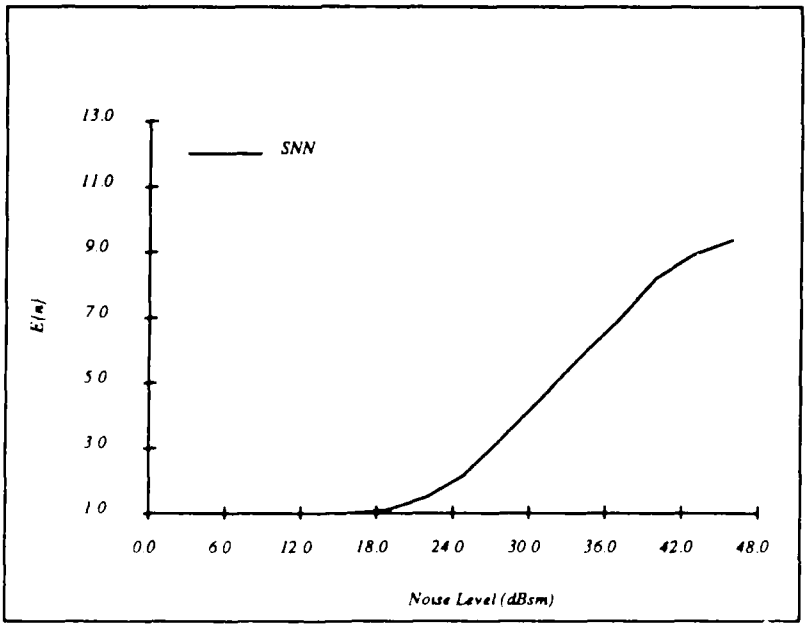


Figure 49: Average number of measurements for the sequential nearest neighbor test with $M = 5$ classes, $N_s = 4$ prototypes/class, $K = 2$.

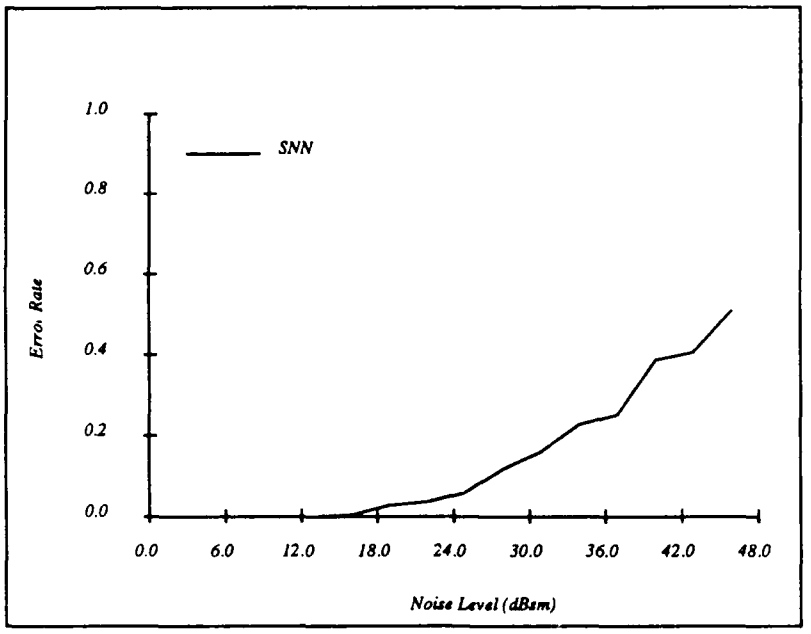


Figure 50: Error rates for the sequential nearest neighbor test with $M = 5$ classes, $N_s = 1$ prototypes/class, $K = 1$.

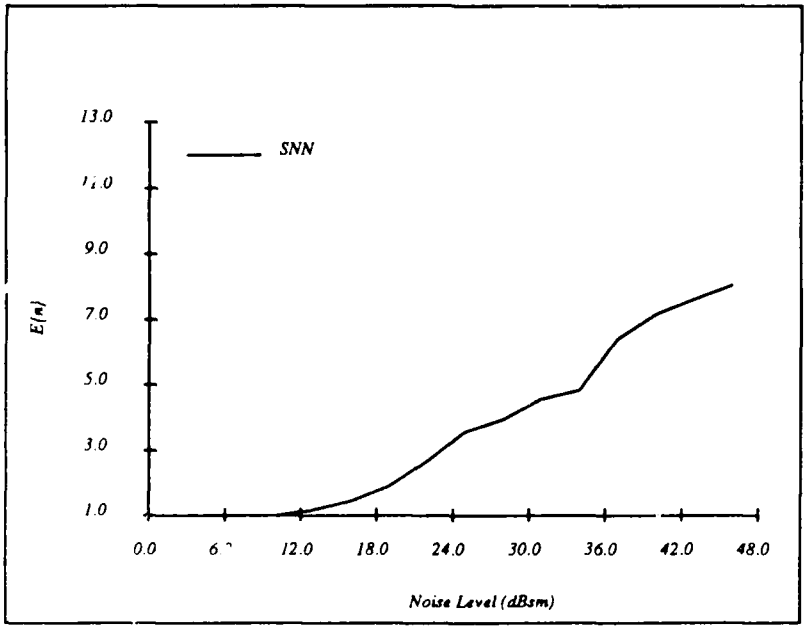


Figure 51: Average number of measurements for the sequential nearest neighbor test with $M = 5$ classes, $N_s = 1$ prototypes/class, $K = 1$.

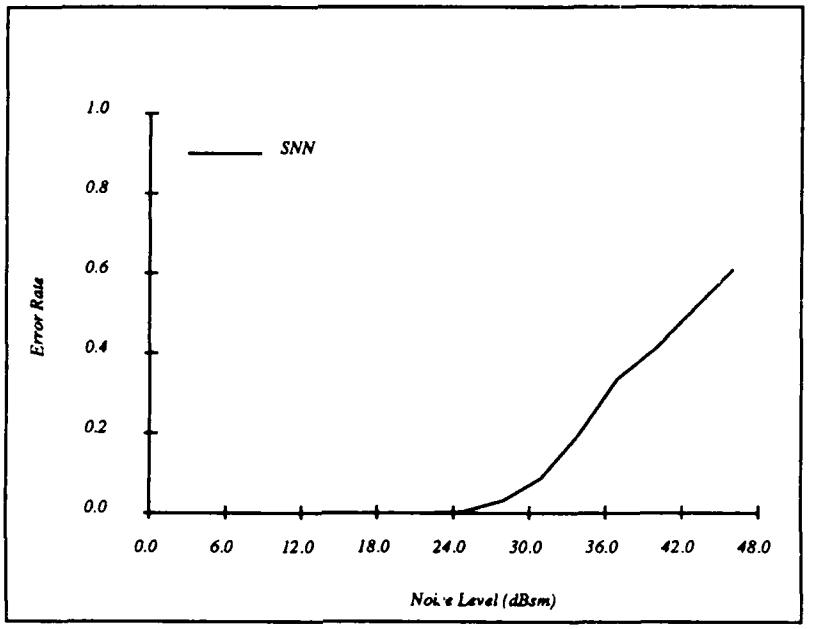


Figure 52: Error rates for the sequential nearest neighbor test with $M = 5$ classes, $N_s = 2$ prototypes/class, $K = 2$.

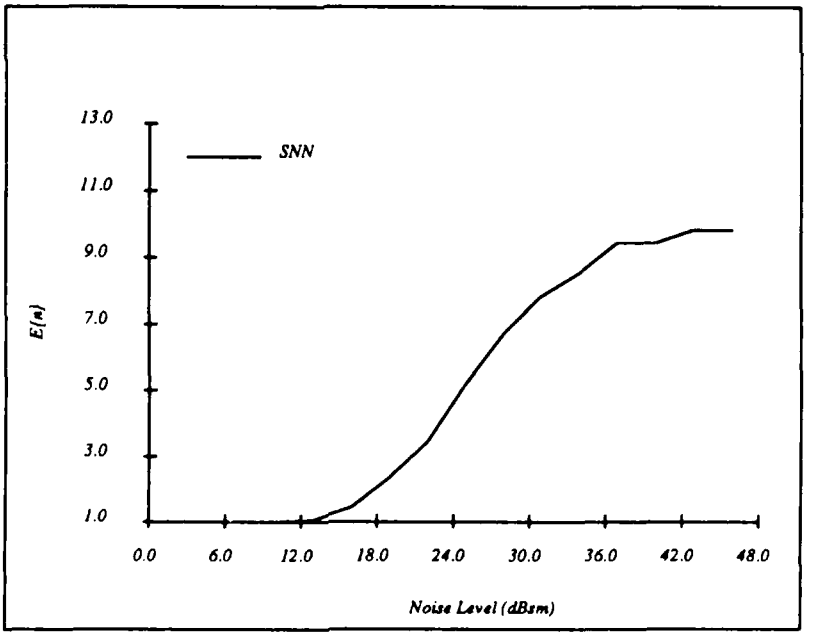


Figure 53: Average number of measurements for the sequential nearest neighbor test with $M = 5$ classes, $N_s = 2$ prototypes/class, $K = 2$.

Table 9: Error rate and average number of measurements for both the sequential nearest neighbor technique and the fixed nearest neighbor technique with $M = 5$ classes, $N_s = 1$ prototypes/class, and $K = 4$.

Noise (dBsm)	Error (SNN)	$E\{n\}(SNN)$	$Error \quad NN \quad N = 10$
30	0.028	7.98	0.012
32	0.044	8.69	0.048
34	0.083	9.22	0.056
36	0.130	9.69	0.120
38	0.192	9.80	0.200

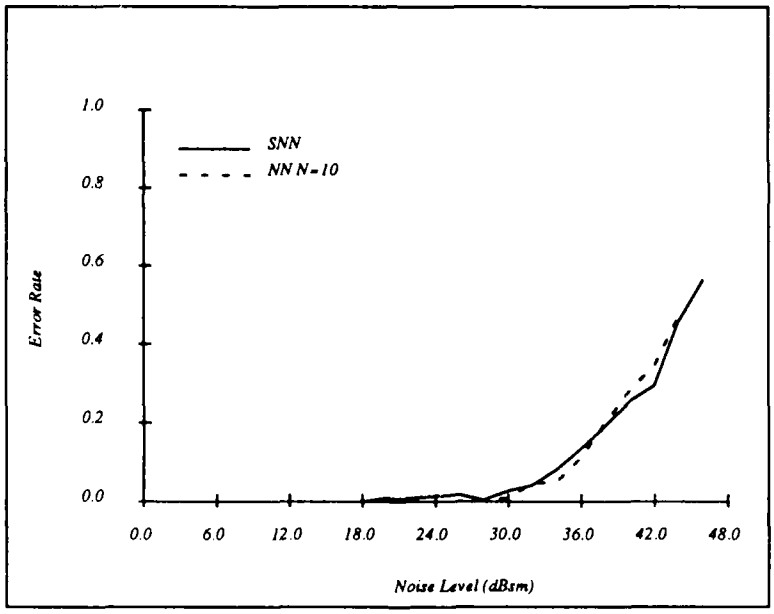


Figure 54: Comparison of the error rates for the sequential nearest neighbor test and the fixed nearest neighbor test with $M = 5$ classes, $N_s = 1$ prototypes/class, $K = 4$.

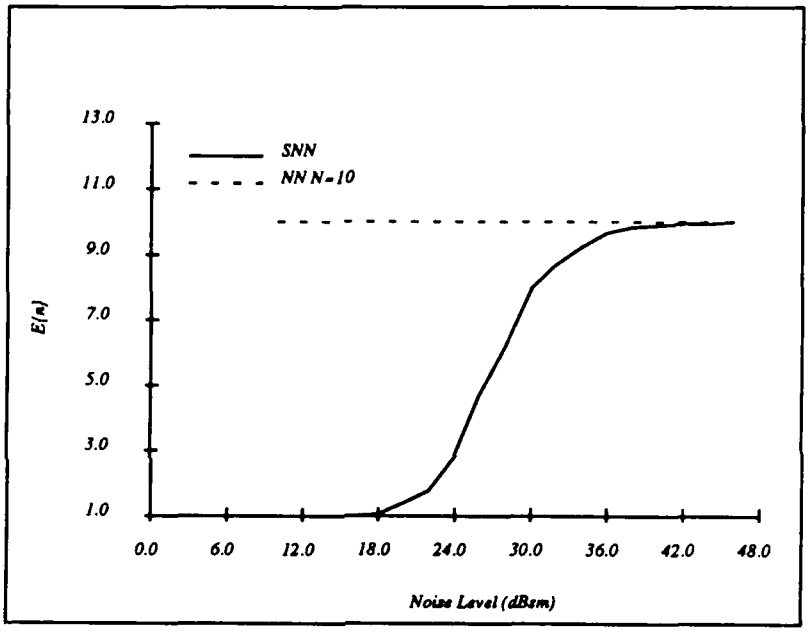


Figure 55: Comparison of the average number of measurements for the sequential nearest neighbor test and the fixed nearest neighbor test with $M = 5$ classes, $N_s = 1$ prototypes/class, $K = 4$.

bor with fixed number of observations is shown in Figures 54-55. Figure 54, shows that the probability of error is almost the same for both algorithms, while Figure 55 shows the reduction in the average number of measurements due to the sequential nearest neighbor. Thus, Figure 55, can be considered as a plot of the efficiency of the sequential nearest neighbor. Tabulated results of this comparison are shown in Table 9.

5.4 Comparison

A number of conclusions can be drawn from these results. First, notice that the sequential nearest neighbor test produces a reasonably low error rate, but requires a large average number of measurements because of its nonparametric nature. In addition, at high noise power levels, ($>$ 20 dBsm) this test frequently requires the maximum number of measurements ($N = 10$) before reaching a decision. However, this test requires only the computation of $M \times N_s$ vector distances at each stage of the sequential test. Moreover, the variation in the error rate with respect to the noise power level is almost linear. Thus, at very high noise power levels, the sequential nearest neighbor gives low error rate compared with the other parametric algorithms.

The Armitage technique also produces low error rates at lower noise power levels but the average number of measurements for this test seems to be particularly sensitive to the noise level, and it requires a large number of measurements for levels exceeding 20 dBsm. The reason is that pairwise comparison is very restrictive, especially at high noise power levels. In addition, this algorithm is relatively complex because it requires the computation of $M(M-1)/2$ likelihood ratios at each stage of the sequential test. However, as shown in Figures 58-63, this test provides the best (among the presented algorithms) compromise of the tradeoff

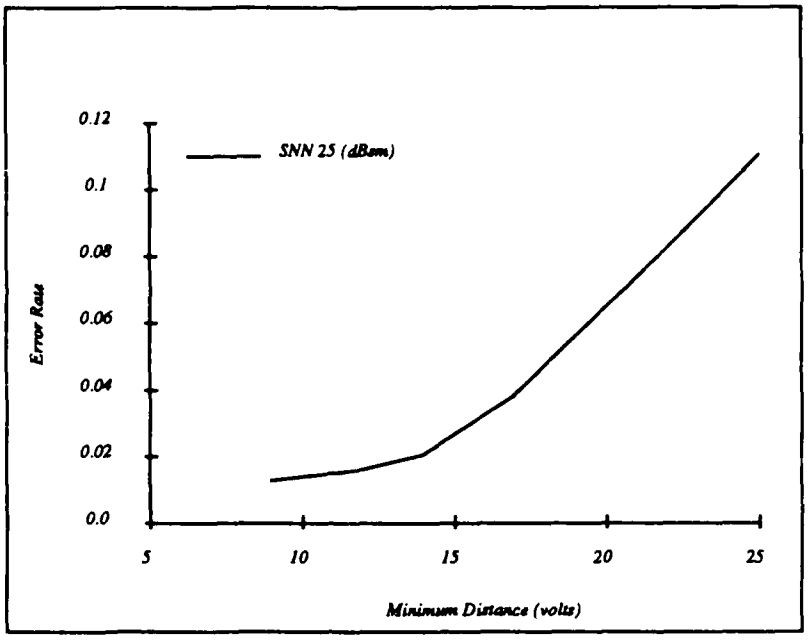


Figure 56: Error rates for the sequential nearest neighbor test as a function of the minimum distance at a 25 (dBsm) noise level with $M = 5$ classes, $N_s = 6$ prototypes/class, $K = 4$.

between the average number of observation and the probability of error.

The method due to Palmer requires the computation of single likelihood ratio at each stage of the test. This method gives low error rate, compared to the other algorithms, at low noise power levels (< 15 dBsm). However, the performance of the Palmer test is not satisfactory at high noise power levels (> 20 dBsm) since it gives higher error rate than other algorithms with almost the same number of observations. The reason for this is that a decision based on comparison of the two largest likelihood functions is not reliable at high noise power levels because all of these functions approach each other at such noise levels.

The Reed algorithm (geometric mean comparison) requires the computation of M likelihood ratios only. This algorithm does not provide a good compromise of the tradeoff between the error probability and the average number of measurements

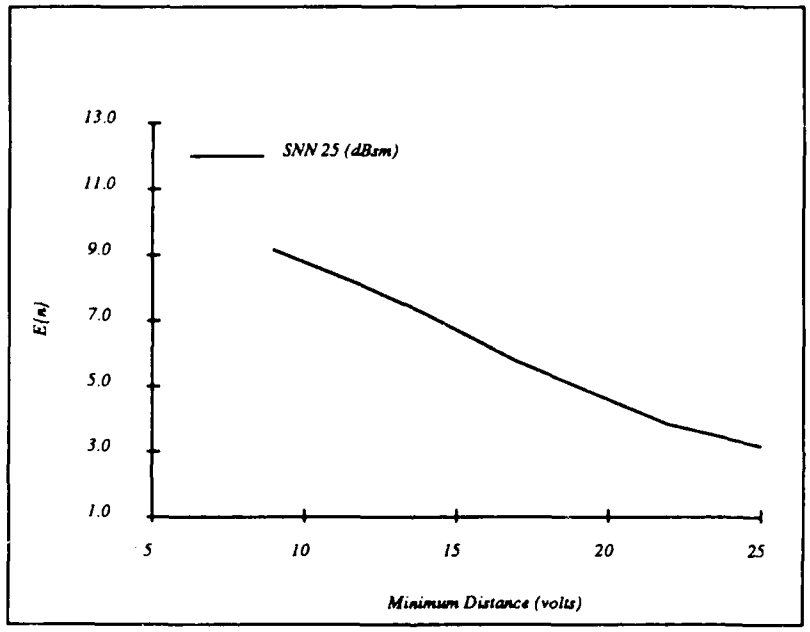


Figure 57: Average number of measurements for the sequential nearest neighbor test as a function of the minimum distance at a 25 (dBsm) noise level with $M = 5$ classes, $N_s = 6$ prototypes/class, $K = 4$.

especially at low noise power levels. The reason is that at low noise levels all likelihood functions have small values except for that of the unclasssified target and the geometric mean of all of the likelihood functions does not carefully represent the likelihood function of the true hypothesis. However, this algorithm is less sensitive to noise power levels than the other parametric algorithms in that the average number of observations increases smoothly as the noise power increases.

Finally, Figures 42-43 imply that the best performance is attained by applying the tree method suggested in this study. While this test requires more computations than most of the other tests, it generally requires the least number of observations and the error compares favorably to that of the other tests. The tree test gives a good compromise of the tradeoff between the number of oservations and the probability of error. However the performance of this test depends on the number of classes and the types of thresholds chosen. In addition, this algorithm is not as complex as the Armitage test since it requires the computation of one likelihood ratio at each stage, and even fewer computations are required as the test progresses until each likelihood function represents a single hypothesis as in Wald test. Moreover, it is possible to predict the average number of observations required by the tree algorithm since the number of observations is predictable at each level of the algorithm.

The sequential maximum *a posteriori* algorithm proposed in this study, which is a modification to the Reed test, reduces the number of observation required by the Reed algorithm without altering the error rate and also without increasing the complexity of the Reed algorithm. In addition, this test can be considered as a suboptimal solution to the Bayes sequential test.

5.5 Sequential Classifier Operating Characteristics (SCOC)

The performance of each of the sequential classification algorithms discussed in this study is evaluated according to its operating characteristics. In Figures 58-63, the probability of misclassification is plotted against the average number of observations for each of the sequential classification algorithms. These curves are considered as operating characteristics for the sequential classifiers. Moreover, these curves can be used for comparison of the various techniques discussed in this study. Thus, a sequential classifier can be designed, using these operating characteristics, such that it will require a certain number of observations for an accepted probability of error.

Figure 63 shows that for the modified version of the Armitage threshold proposed in this study, the best reduction in the average number of observations is attained when $r = 1$.

Figures 58-63, show that the Armitage and the tree tests are the best in compromising the tradeoff between the probability of error and the average number of measurements. In addition, while the Armitage technique requires less observations than the tree method for high error probabilities (> 3 percent), the tree technique requires fewer observations for very low probability of error than the Armitage technique. Moreover, it is clear from these figures that there is a reduction in the average number of observations needed by sequential methods compared to the maximum a posteriori test (Likelihood) with fixed number of observations. This reduction reaches 60 percent at some noise levels and for certain error probabilities. However, for high error probability (> 10 percent), the operating characteristics of all the sequential classification techniques approach each other. Finally, it is clear that the performance of each of the sequential algorithms

depends on the computations required and the complexity of the sequential test.

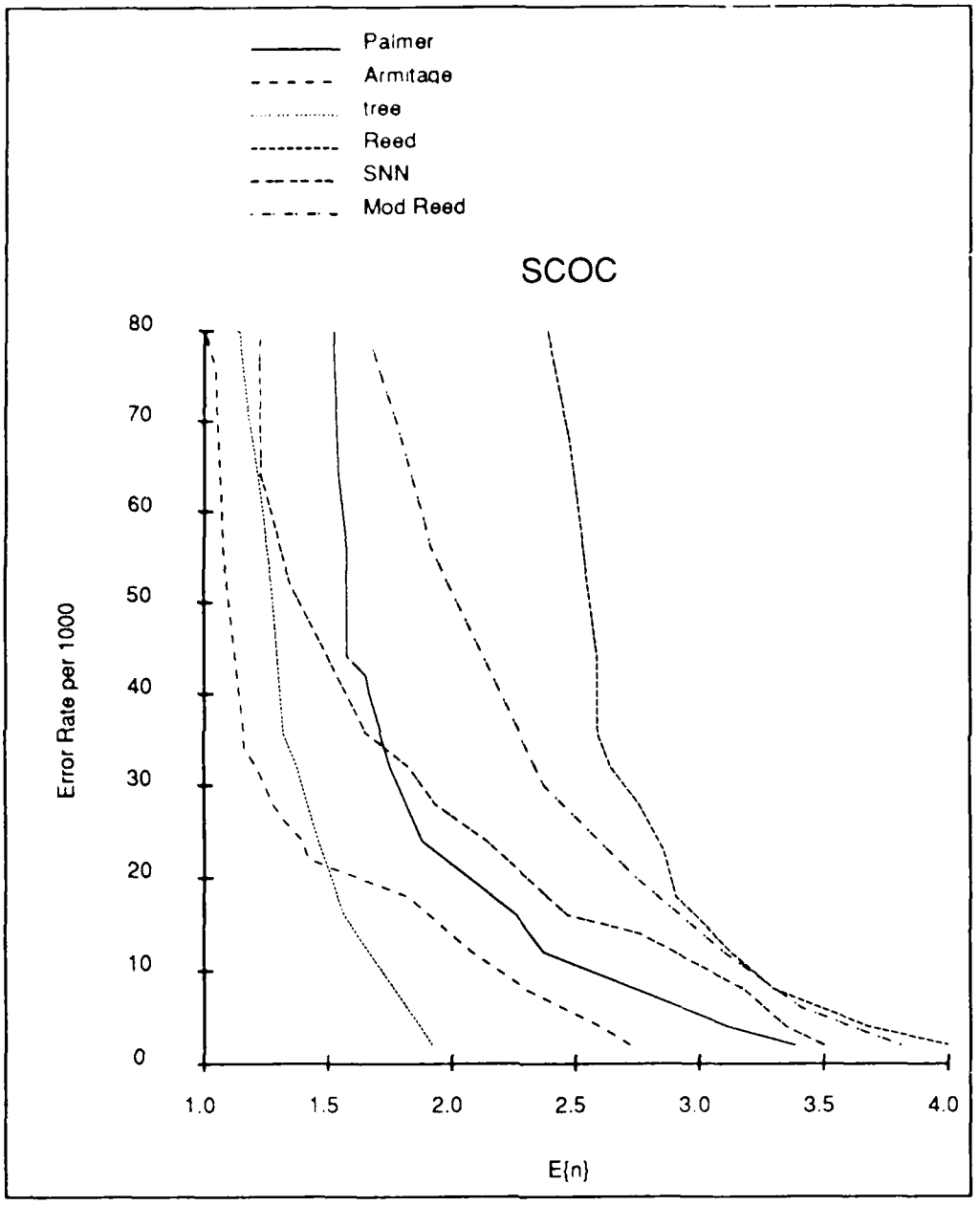


Figure 58: Comparison of the sequential classifiers operating characteristics at 25 (dBsm) noise level with $M = 5$, $N_s = 1$ prototypes/class, $K = 4$.

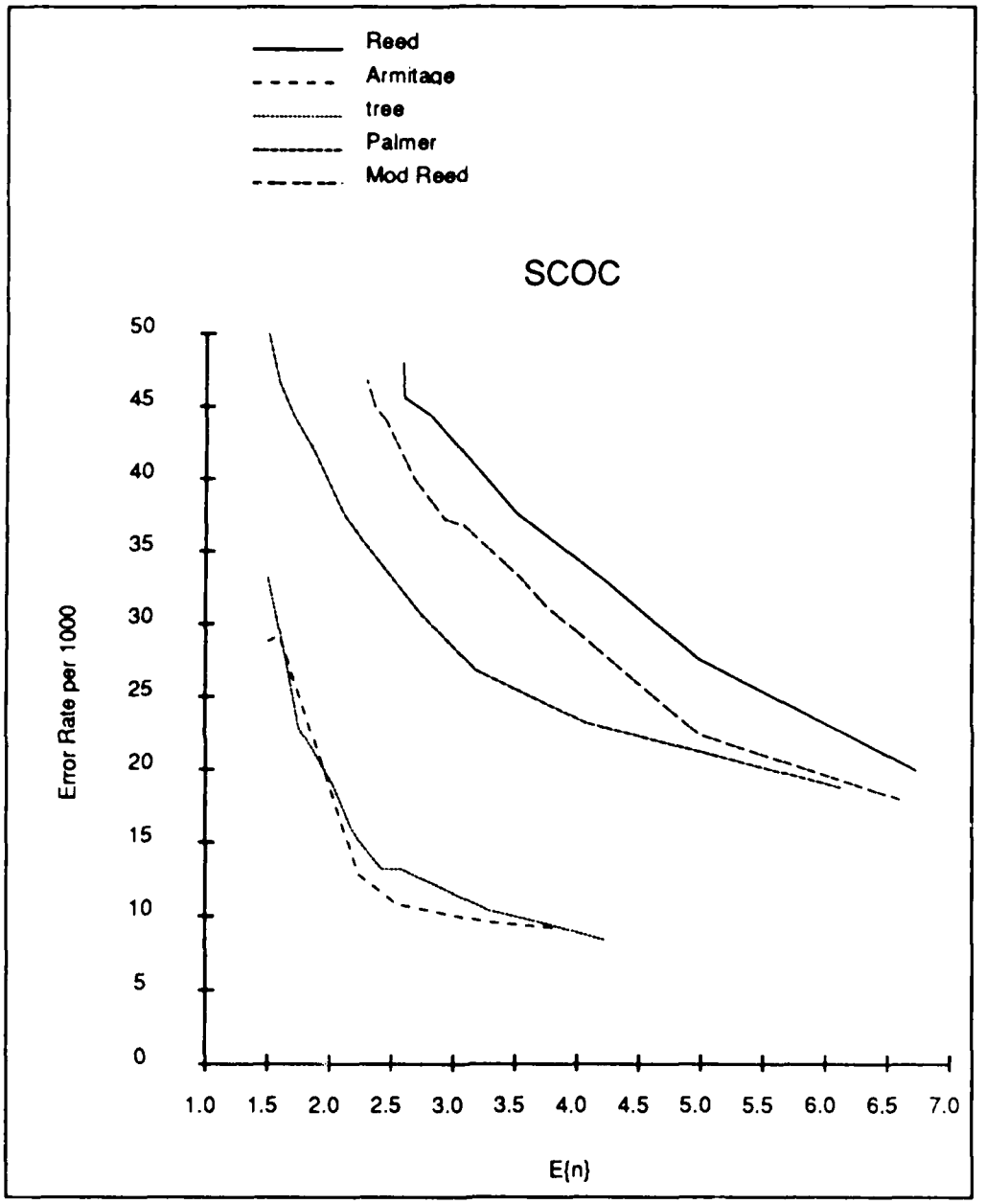


Figure 59: Comparison of the sequential classifiers operating characteristics at 30 (dBsm) noise level with $M = 5$, $N_s = 1$ prototypes/class, $K = 4$.

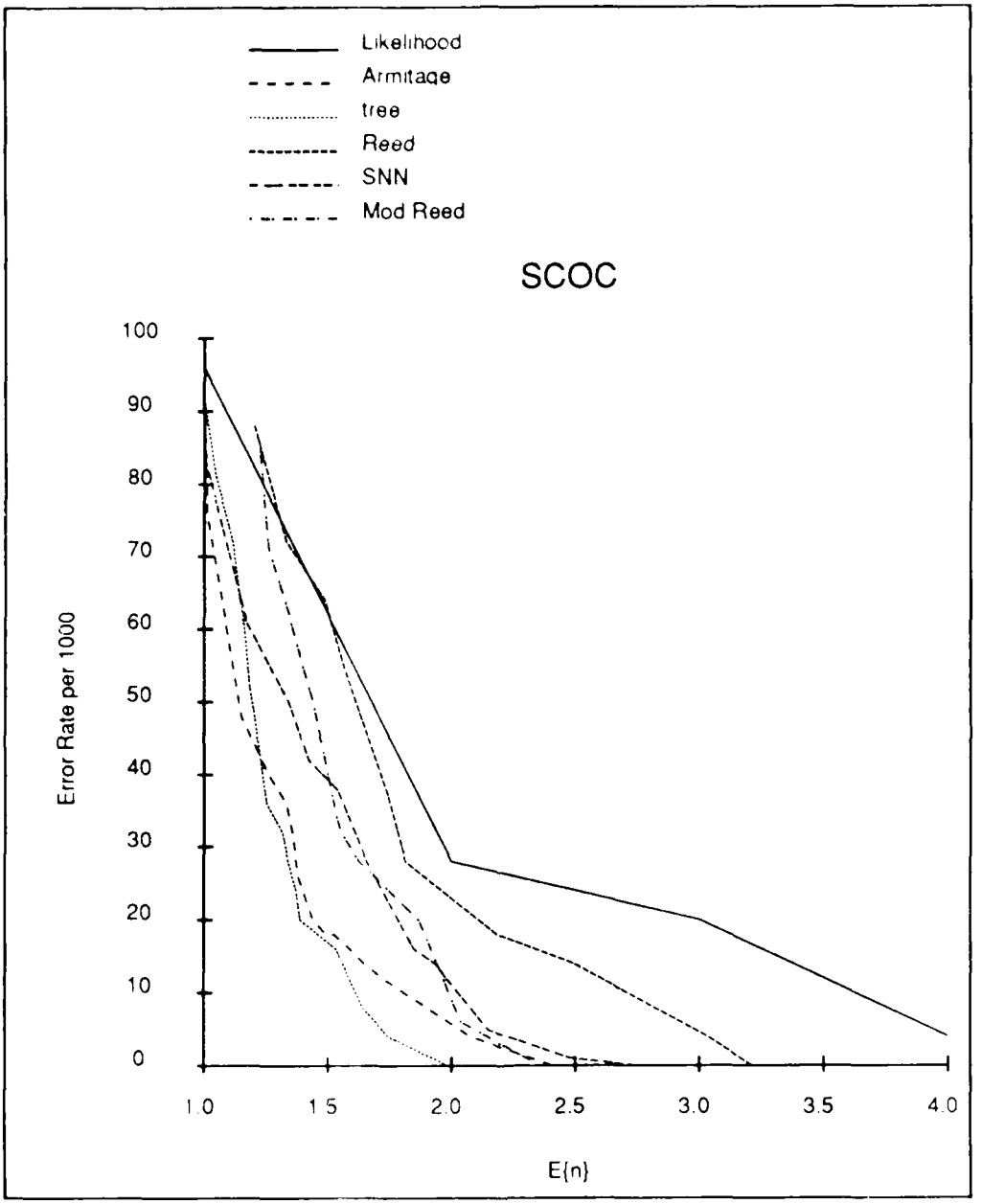


Figure 60: Comparison of the sequential classifiers operating characteristics at 20 (dBsm) noise level with $M = 5$, $N_s = 2$ prototypes/class, $K = 2$.

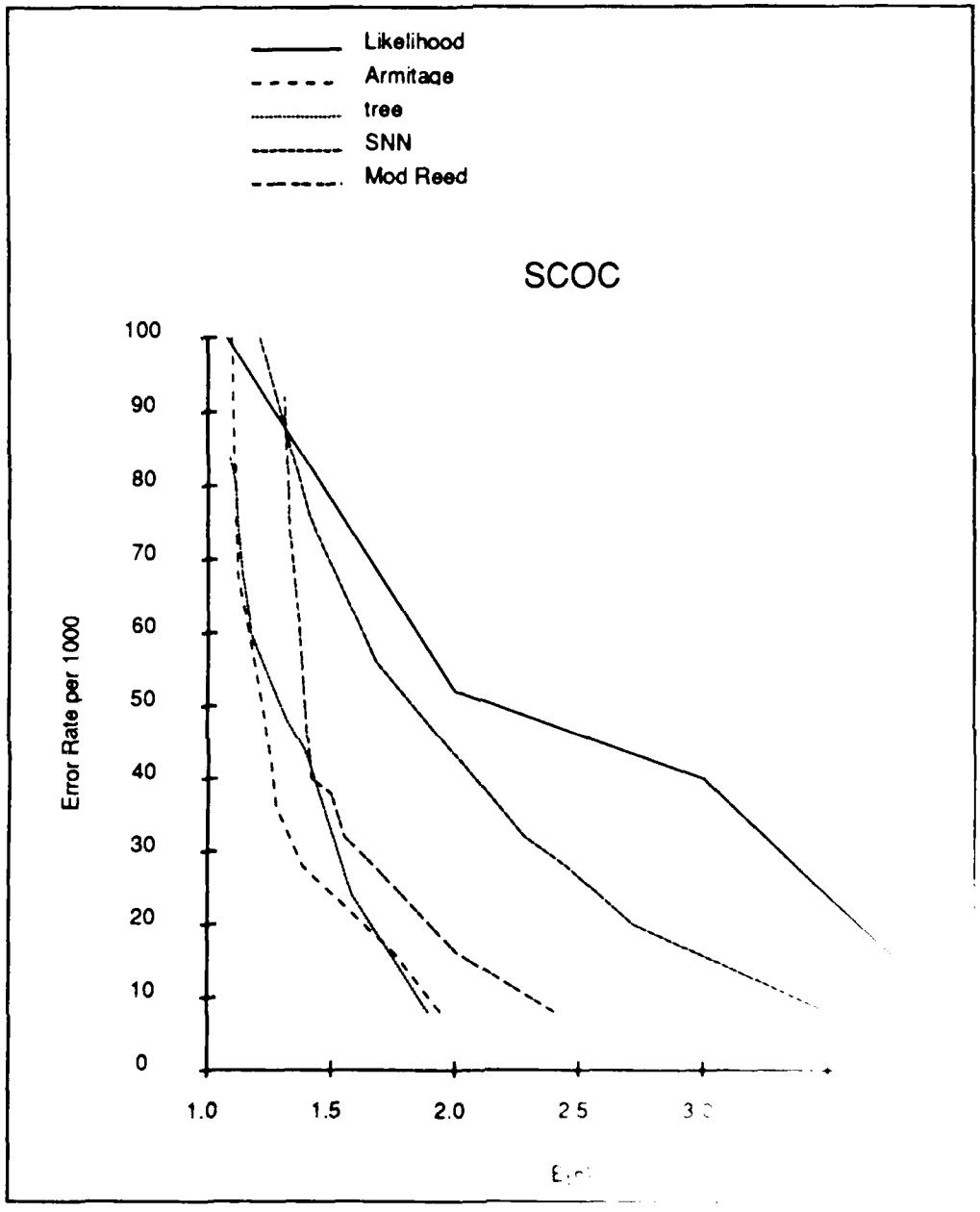
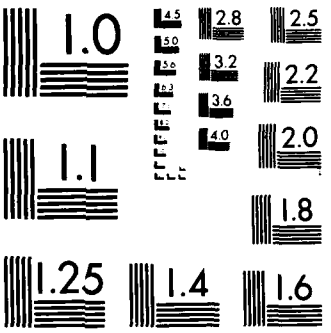


Figure 61: Comparison of the sequential [illegible]
(dBsm) noise level with M [illegible]

AD-A184 894 ON M-ARY SEQUENTIAL HYPOTHESES TESTING FOR THE CLASSIFICATION OF RADAR SIGNALS(U) OHIO STATE UNIV COLUMBUS ELECTROSCIENCE LAB I JOUNY ET AL JUL 87 2/2

UNCLASSIFIED ESL-717220-4 N00014-85-K-0321 F/G 17/9 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

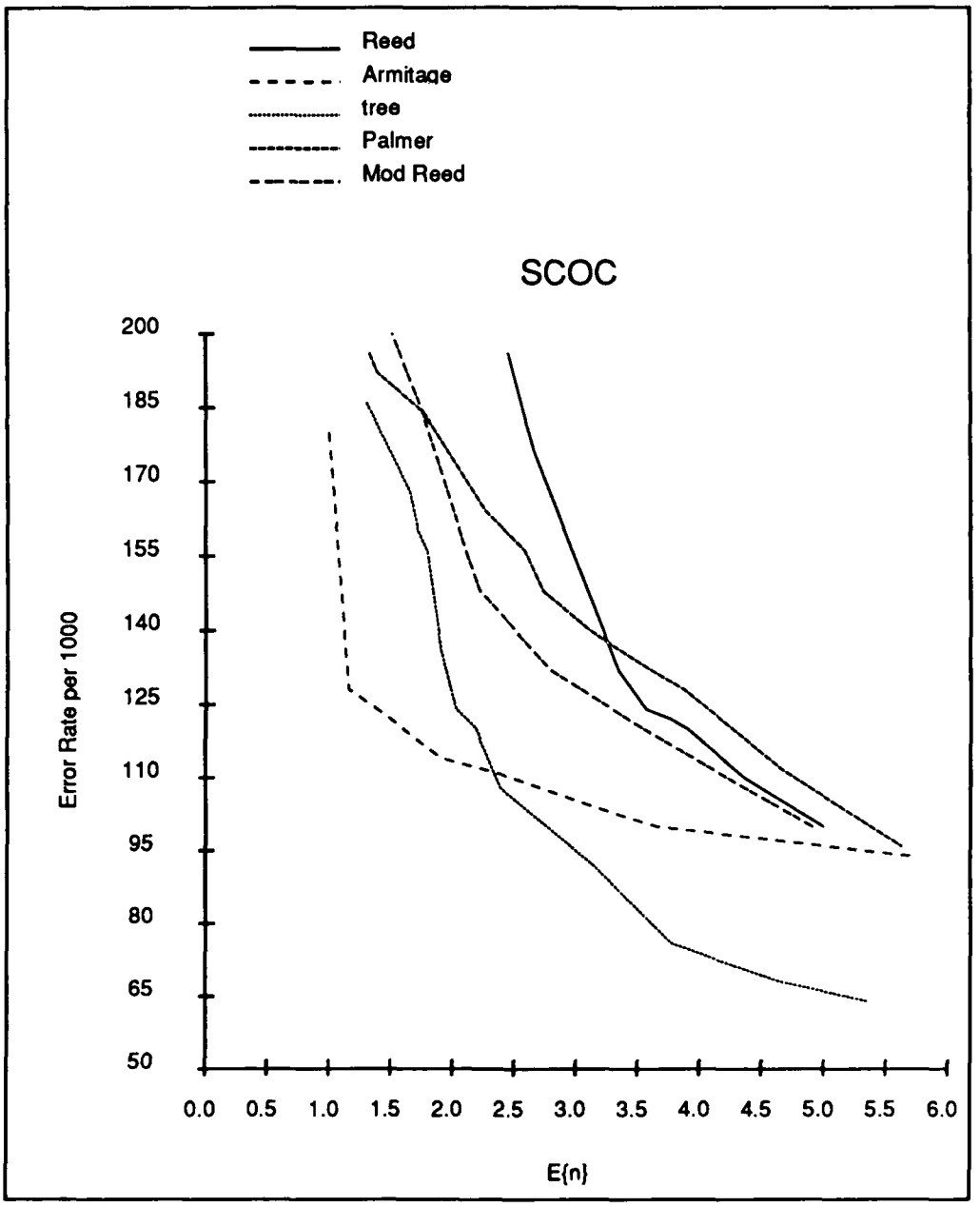


Figure 62: Comparison of the sequential classifiers operating characteristics at 25 (dBsm) noise level with $M = 5$, $N_s = 2$ prototypes/class, $K = 2$.

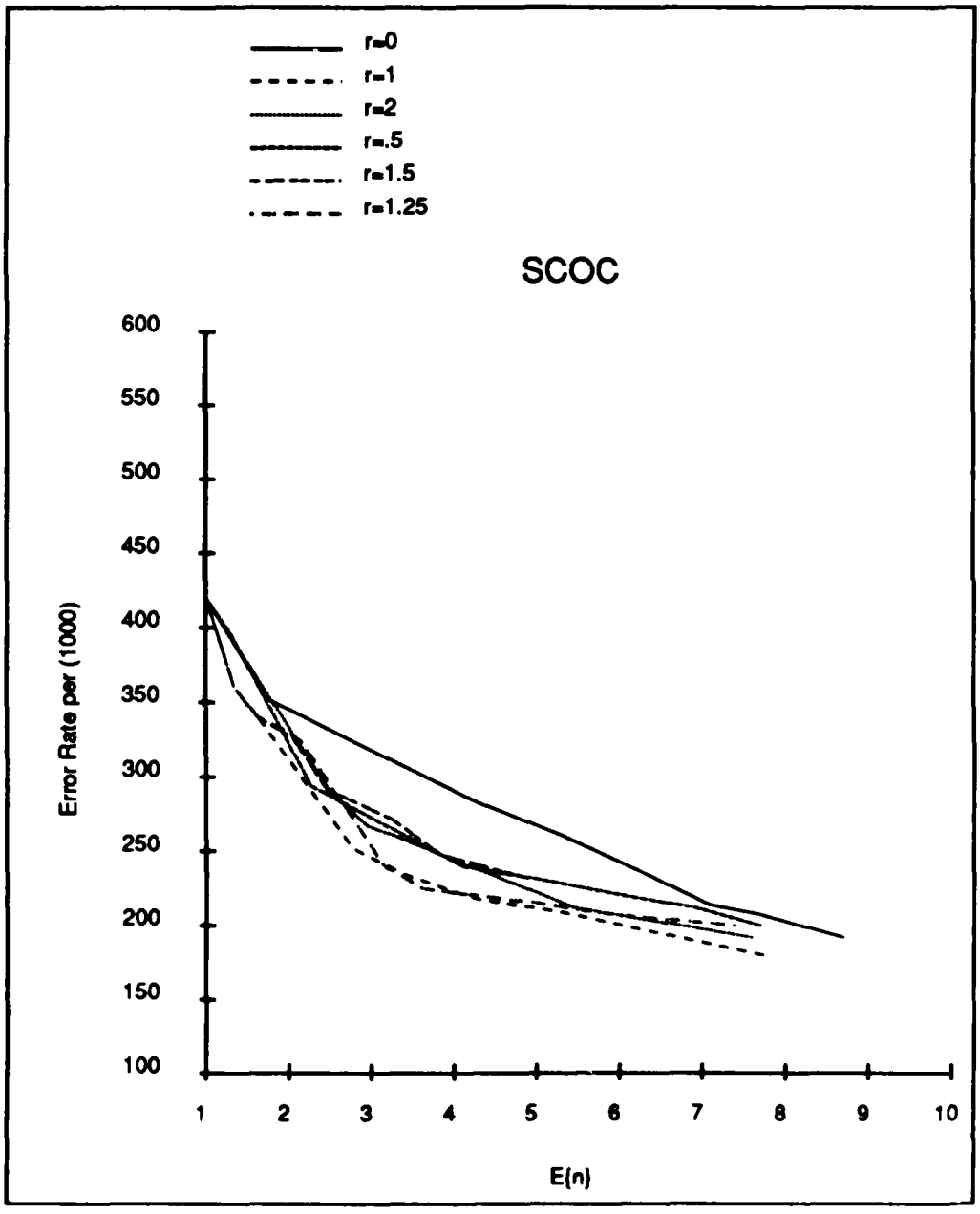


Figure 63: Comparison of the sequential classifier operating characteristics for the Armitage technique at 30 (dBsm) noise level with $M = 5$, $N_s = 2$ prototypes/class, $K = 2$.

CHAPTER VI

Conclusions

The simulation results presented in Chapter V indicate that sequential hypothesis testing techniques may realize significant advantages for applications to radar target identification problems. The flexibility of these techniques and their ability to satisfy practical constraints on the classification error performance of radar systems make the incorporation of sequential algorithms particularly attractive for RTI.

Sequential methods can be used to tune and direct interrogating radar for maximum reliability. Moreover, sequential hypotheses techniques can announce early, low-confidence decisions and, as more data is aquired, high confidence decisions. Thus, sequential techniques are flexible.

By changing the parameters that define the decision regions, we can alter the performance of the test to match our bounds on error probability and test length (number of observations). A common feature of all the parametric and nonparametric sequential algorithms presented in this study is that we can alter the performance of these techniques by modifying the decision boundaries to make them dependent on the number of observations.

Some of the M ary hypotheses techniques with fixed number of observations can be modified into sequential techniques by specifying a certain null region in the hypotheses decision space. Thus, the number of observations can be reduced while

the probability of error remains unchanged. The reason is that when observations are taken sequentially, the classification test might terminate before requesting all observations specified by a fixed number of measurements test.

The performance of the various sequential classification algorithms depends on the noise power level. Thus, we may choose to use more than one classification algorithm at different noise power levels provided that an *a priori* knowledge of the noise power level is available. For example, we may use the method due to Palmer at low noise power levels because of its simplicity and also the tree algorithm at high noise power levels.

It is clear from this study that the performance of any sequential classification algorithm is dependent on the complexity of the algorithm. In general, the tree algorithm proposed in this study, minimizes the average number of observations, with less complexity than the pairwise comparison (the Armitage test).

Truncation is a major factor in sequential classification algorithms because the number of observations is usually finite. Thus, an optimal form of truncation is necessary to achieve a good performance. It is also clear that the optimization of parameters as the maximum number of measurements, and the choice of the functional form of the various decision thresholds deserves further investigation.

6.1 Noise Dependent Group Sequential Tests

One common feature of all of the sequential techniques discussed in this study is that a large number of observations is required at high noise levels. This is clearly shown in Chapter V. Whenever an observation is repeated, hypotheses testing is performed at each stage of the sequential test.

If *a priori* knowledge about the noise power level is available, which is the case for parametric techniques, then the following modifications can be applied to

any of the sequential tests discussed.

1. The maximum allowed number of observations N is chosen as a function of the noise power σ^2. That is, N is proportional to the noise power.

2. Groups of observations, rather than single observation, are repeated one at a time. observation. The size of these groups is dependent on both the noise power and the stage of the test. For example if a group of 4 measurements is observed at the begining, the size of the second group could be 2 or 1.

It is obvious that the above modification reduces the complexity of the sequential classification algorithms because hypotheses testing is performed after requesting a group of observations rather than single observation at each stage of the sequential test.

REFERENCES

[1] A. Kamis, F. D. Garber, and E. K. Walton, "Radar target classification studies - software development and documentation," Technical Report 716559-1, The Ohio State University, Department of Electrical Engineering, ElectroScience Laboratory, September 1985.

[2] A. A. Ksienski, Y. T. Lin, and L. J. White, "Low-frequency approach to target identification," *Proceedings of the IEEE*, vol. 63, no. 12, pp. 1651–1660, December 1975.

[3] A. Wald, *Sequential Analysis*. New York: Wiley, 1947.

[4] K. S. Fu, *Sequential Methods in Pattern Recognition and Machine Learning*. New York: Academic Press, 1968.

[5] A. Wald and J. Wolfowitz, "Optimum character of the sequential probability ratio test," *Ann. Math. Stat.*, vol. 19, pp. 326–339, 1948.

[6] P. Armitage, "Sequential analysis with more than two alternative hypotheses, and its relation to discriminant function analysis," *Journal of the Royal Statistical Society*, Ser. B, vol. 12, pp. 137–144, November 1950.

[7] R. C. Reed, "A sequential multi-decision procedure," in *Proceedings of the Symposium on Decision Theory and Applications to Electronic Equipment Development*, USAF Development Center, Rome NY, April 1960.

[8] S. E. Smith and S. S. Yau, "Linear sequential pattern classification," *IEEE Transactions on Information Theory*, vol. IT-18, no. 5, pp. 673-678, September 1972.

[9] E. K. Walton and J. D. Young, "The Ohio State University compact radar cross-section measurement range," *IEEE Transactions on Antennas and Propagation*, vol. AP-32, no. 11, pp. 1218-1223, November 1984.

[10] Ögmundur Snorrason, "Feature selection applied to radar target identification," M.S. Thesis, The Ohio State University, June 1987. also: ElectroScience Laboratory Report 717220-3.

[11] H. L. Van Trees, *Detection, Estimation, and Modulation Theory: Part I.* New York: Wiley, 1968.

[12] C. C. Lee and J. B. Thomas, "A modified sequential detection procedure," *IEEE Transactions on Information Theory*, vol. IT-30, no. 1, pp. 16-23, January 1984.

[13] S. Tantaratana and J. B. Thomas, "Relative efficiency of the sequential probability ratio test in signal detection," *IEEE Transactions on Information Theory*, vol. IT 24, no. 1, pp. 22-31, January 1978.

[14] C. H. Chen, *Statistical Pattern Recognition.* Rochelle Park, N. J.: Hayden Book Co., 1973.

[15] S. Fleisher and E. Shwedyk, "A sequential multiple hypothesis test for the unknown parameters of a Gaussian distribution," *IEEE Transactions on Information Theory*, vol. IT 26, no. 2, pp. 255 259, March 1980.

[16] L. C. Palmer, "Sequential tests to select among m hypotheses," *IEEE Transactions on Information Theory*, vol. IT–18, no. 1, pp. 211–214, January 1972.

[17] S. N. Srihari, "Comparitive evaluation of the Sebestyen and nearest neighbor classifiers for invariant pattern recognition," Ph.D. Dissertation, The Ohio State University, June 1976.

[18] J. T. Tou and R. C. Gonzalez, *Pattern Recognition Princilpes*. Reading, Massachusetts: Addison–Wesley, 1974.

[19] T. Y. Young and K. Fu, Eds., *Hanbook of Pattern Recognition and Image Processing*. Orlando: Academic Press, 1986.

END

10-87

DTIC